MEDINFO 2015

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The series has been accepted by MEDLINE/PubMed, SciVerse Scopus, EMCare, Book Citation Index – Science and Thomson Reuters' Conference Proceedings Citation Index.

Volume 216

Recently published in this series

Vol. 215. L. Botin, P. Bertelsen and C. Nøhr (Eds.), Techno-Anthropology in Health In-formatics – Methodologies for Improving Human-Technology Relations
Vol. 214. A. Georgiou, H. Grain and L.K. Schaper (Eds.), Driving Reform: Digital Health is Everyone's Business – Selected Papers from the 23rd Australian National Health Informatics Conference (HIC 2015)
Vol. 213. J. Mantas, A. Hasman and M.S. Househ (Eds.), Enabling Health Informatics Applications
Vol. 212. D. Hayn, G. Schreier, E. Ammenwerth and A. Hörbst (Eds.), eHealth2015 – Health Informatics Meets eHealth
Vol. 211. B. Blobel, M. Lindén and M.U. Ahmed (Eds.), pHealth 2015 – Proceedings of the 12th International Conference on Wearable Micro and Nano Technologies for Personalized Health, 2–4 June 2015, Västerås, Sweden
Vol. 210. R. Cornet, S.K. Andersen, L. Stoicu-Tivadar, C.L. Parra Calderón, M. Hercigonja-Szekeres and A. Hörbst (Eds.), Digital Healthcare Empowering Europeans: Proceedings of MIE2015
Vol. 209. G. Gillis, D. Newsham and A.J. Maeder (Eds.), Global Telehealth 2015: Integrating Technology and Information for Better Healthcare
Vol. 208. K.L. Courtney, A. Kuo and O. Shabestari (Eds.), Driving Quality in Informatics: Fulfilling the Promise
Vol. 207. M. Graña, C. Toro, R.J. Howlett and L.C. Jain (Eds.), Innovation in Medicine and Healthcare 2014
Vol. 206. A.J. Maeder, M. Mars and R.E. Scott (Eds.), Global Telehealth 2014
Vol. 205. C. Lovis, B. Séroussi, A. Hasman, L. Pape-Haugaard, O. Saka and S.K. Andersen (Eds.), e-Health – For Continuity of Care: Proceedings of MIE2014
Vol. 204. H. Grain, F. Martin-Sanchez and L.K. Schaper (Eds.), Investing in E-Health: People, Knowledge and Technology for a Healthy Future – Selected Papers from the 22nd Australian National Health Informatics Conference (HIC 2014)
Vol. 203. G. Riva, P. Ajmone Marsan and C. Grassi (Eds.), Active Ageing and Healthy Living – A Human Centered Approach in Research and Innovation as Source of Quality of Life

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

MEDINFO 2015: eHealth-enabled Health

Proceedings of the 15th World Congress on Health and Biomedical Informatics

Edited by

Indra Neil Sarkar

Brown University

Andrew Georgiou

Macquarie University

and

Paulo Mazzoncini de Azevedo Marques

University of São Paulo

IOS Press

Amsterdam • Berlin • Tokyo • Washington, DC

ISBN 978-1-61499-563-0 (print)
ISBN 978-1-61499-564-7 (online)

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

Introductory Remarks from the Scientific Program Chairs

Fernando Martin-Sanchez[a], Kaija Saranto[b]

[a] *Health and Biomedical Informatics Centre, Melbourne Medical School, The University of Melbourne, Australia*
[b] *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*

MEDINFO is the premier international Health and Biomedical Informatics event. MEDINFO 2015 is hosted by SBIS (Brazilian Health Informatics Society) on behalf of the International Medical Informatics Association (IMIA) and will take place in the city of São Paulo from the 19th to 23rd August 2015. MEDINFO 2015 continues a 41-year tradition of bringing together world leaders, policy makers, researchers, practitioners, educators, and students to exchange ideas and contribute to the latest developments, innovations, and global trends in this rapidly advancing, multidisciplinary field of Health and Biomedical Informatics. This is the first MEDINFO that has been organized to reflect the new two-yearly cycle approved by IMIA. We were thus very happy when we reached the submission and registration deadlines with numbers very similar to previous MEDINFOs that had been organized in three-yearly cycles.

Under the theme: "*eHealth-enabled Health*", the world leaders in this field will gather in Brazil to share knowledge and analyze how Health and Biomedical Informatics is contributing to address some of the most challenging problems in health care, public health, consumer health and biomedical research. Researchers, clinicians, technologists and managers will attend and share experiences on the use of information methods, systems and technologies to promote patient-centered care, improve patient safety, enhance care outcomes, facilitate translational research, enable precision medicine and improve education and skills in health informatics.

This is an historical event as MEDINFO is hosted in Latin America for the first time. Inclusiveness has been a main goal in MEDINFO 2015 with affordable registration fees for the regional audience and use of Spanish and Portuguese language in tutorials and simultaneous translation in sessions held in the main auditorium. MEDINFO 2015 features a pre-congress offering of an extensive tutorial program by leading experts and a student paper competition that draws the best young talent from all over the world. The main program includes keynote talks, papers, posters, panels, workshops, and scientific demonstrations that span a broad range of topics from emerging methodologies that contribute to the conceptual and scientific foundations of Health and Biomedical Informatics, to successful implementations of innovative application, integration, and evaluation of eHealth systems and solutions.

The conference program features five keynote presentations, 178 paper presentations, 248 poster abstract presentations, 27 panels, 30 workshops and 17 scientific demonstrations.

The contributions and presentations included in the program were carefully selected through a rigorous review process involving almost 400 reviewers for a large number of submissions (793) sent by 2500 authors from 59 countries all over the world. The Scientific Program Committee Co-Chairs are grateful to the four Track Chairs, the members of the Scientific Program Committee and all the reviewers who have contributed to the process, and thank the Editorial Committee, the Local Organizing Committee and the IMIA officers (in particular CEOs and VP Medinfo) for assisting us in putting this program together.

The conference participants come to São Paulo from all continents and 60 different countries. We hope that you will enjoy the published proceedings and the overall program!

Sincerely,

Fernando Martin-Sanchez, PhD, FACHI, FACMI &
Kaija Saranto, PhD, FACMI, FAAN
Co-Chairs, MEDINFO 2015 Scientific Program Committee

This page intentionally left blank

Introductory Remarks from the Editorial Committee Chair

Indra Neil Sarkar[a]

[a] *Center for Biomedical Informatics, Warren Alpert Medical School of Brown University, Providence, RI, USA*

Let me join the rest of the organizing committees in welcoming you to MEDINFO 2015 in São Paulo. As the Editorial Committee Chair, I had the distinct honor to review every accepted submission to this year's congress. I personally wish to extend a thanks to the authors for their fine contributions. Together with the meeting participants, MEDINFO 2015 is positioned to be an unprecedented exposition of the finest biomedical informatics innovations with global impact.

Appreciating the international scope of the MEDINFO congresses, it is essential to embrace principles to support scientific inclusivity. Therefore, in contrast to many scientific meetings, the general criteria used for selection into the MEDINFO proceedings is based mostly on scientific merit; language issues are not reason alone for a submission to be not selected. The cost of this inclusivity is that each accepted submission must be carefully reviewed and edited to adjust for language that does not impact the scientific contribution. It is important to note that even submissions from native English speakers may require editing due to variance from the required template, typographical errors, or grammatical issues.

Building on the framework developed by Christoph Lehmann for MEDINFO 2013, Assistant Editors (AEs) were recruited from biomedical informatics training programs (Table 1). The minimum criterion for selection as an AE was at least one first author peer-reviewed English publication (ideally in an informatics conference or journal). Poster submissions were reviewed by one AE; paper submissions by two AEs. The edits were then finalized and assembled into the final proceedings that are in front of you now.

It is important for authors to understand the costs associated with the editing and overall production efforts to ensure the MEDINFO proceedings are of the highest quality possible. Following my esteemed colleagues who served as Editorial Committee Chairs for previous MEDINFOs, I make a plea to each of you to consider the work that is involved when aiming to circumvent the standards established by the organizing committees.

Even moreso than in previous MEDINFOs, strict adherance to the template guidelines was deemed an essential criterion for inclusion in the proceedings. Nonetheless, a number of submissions did clear the peer-review process that still required formatting edits to ensure consistency in font size, spacing, and overall style. In some instances, text had to be significantly edited or figures drastically shrunken or eliminated all together to ensure page limits were respected. Even with such edits, a good faith effort was still made for preserving the scientific message of the contributions. I am thankful for the dedication and hard work of 26 AEs that worked, word-by-word, through each submission and made edits that were ultimately vetted and approved by me.

Table 1– Assistant Editors (AEs) for MEDINFO 2015

Assistant Editor	Institution
Samira Y. Ali	Carlow University
Andrew B.L. Berry	University of Washington
Haresh L. Bhatia	Vanderbilt University
Richard Brandt	Texas Tech University
Matthew K. Breitenstein	Mayo Clinic
Jacqueline E. Brixey	University of Texas at El Paso
David Chartash	Indiana University
Perry M. Gee	Dignity Health
Mattias Georgsson	Blekinge Institute of Technology
Anupama Edward Gururaj	University of Texas Health Science Center at Houston
Zhe He	Columbia University
Kate Fultz Hollis	Oregon Health & Sciences University
Silis Y. Jiang	Columbia University
Saeed Mehrabi	Mayo Clinic
Amir Mohammad	Yale University
Tiffany Nicole Moncur	University of South Florida
Shauna Marie Overgaard	University of Minnesota
Jennifer Elizabeth Prey	Columbia University
Lisiane Pruinelli	University of Minnesota
Balaji Polepalli Ramesh	University of Massachusetts
Joseph D. Romano	Columbia University
Charlene Ronquillo	University of British Columbia
Ning Shang	Columbia University
Harry Tunnell	Indiana University-Purdue University Indianapolis
Mary Regina Wysocki	University of Texas Health Science Center at Houston
Rafeek Adeyemi Yusuf	University of Texas Health Science Center at Houston

Finally, I wish to acknowledge the other members of the Editorial Committee (Paulo Mazzoncini de Azevedo Marques and Andrew Georgiou), along with Alvaro Margolis (IMIA Vice President for MEDINFO), Peter Murray (Immediate Past IMIA CEO), Elaine Huesing (Interim IMIA CEO), the leadership of the Local Organizing Committee (Beatriz de Faria Leão and Claudio Giulliano Alves da Costa) and the Scientific Program Committee Co-Chairs (Fernando Martin Sanchez and Kaija Saranto). These proceedings and this meeting are the product of a true team effort– I hope you enjoy MEDINFO 2015 in São Paolo!

Sincerely,

Indra Neil Sarkar, PhD, MLIS, FACMI
Chair, Editorial Committee

This page intentionally left blank

Contents

Introductory Remarks from the Scientific Program Chairs v
Fernando Martin-Sanchez and Kaija Saranto

Introductory Remarks from the Editorial Committee Chair vii
Indra Neil Sarkar

Part 1

Papers

Electronic Health Records

Validating the Access to an Electronic Health Record: Classification and Content Analysis of Access Logs 3
Leandro Noer Alassia, Sonia Benítez, Daniel Roberto Luna and Fernán González Bernaldo de Quiros

A Practical Approach to Governance and Optimization of Structured Data Elements 7
Sarah A. Collins, Emily Gesner, Steven Morgan, Perry Mar, Saverio Maviglia, Doreen Colburn, Diana Tierney and Roberto Rocha

Systems Architecture for a Nationwide Healthcare System 12
Jorge Abin, Horacio Nemeth and Ignacio Friedmann

Electronic Dental Records System Adoption 17
Renata Abramovicz-Finkelsztain, Claudia G. N. Barsottini and Heimar Fatima Marin

Characterizing the Structure of a Patient's Care Team Through Electronic Encounter Data Analysis 21
Shan He, Greg Gurr, Susan Rea and Sidney N. Thornton

An Information Paradigm Shift is Required to Realize EHR Benefits 26
Evelyn Hovenga and Heather Grain

Template and Model Driven Development of Standardized Electronic Health Records 30
Stefan Kropf, Claire Chalopin and Kerstin Denecke

Fast Model Adaptation for Automated Section Classification in Electronic Medical Records 35
Jian Ni, Brian Delaney and Radu Florian

Using EHRs and Machine Learning for Heart Failure Survival Analysis 40
Maryam Panahiazar, Vahid Taslimitehrani, Naveen Pereira and Jyotishman Pathak

Towards the Implementation of an openEHR-Based Open Source EHR Platform (A Vision Paper) 45
Pablo Pazos Gutiérrez

An Ecosystem of Intelligent ICT Tools for Speech-Language Therapy Based on a Formal Knowledge Model 50
Vladimir Robles-Bykbaev, Martín López-Nores, José Pazos-Arias, Diego Quisi-Peralta and Jorge García-Duque

Identification of Patient Safety Risks Associated with Electronic Health Records: A Software Quality Perspective 55
Luiz A. Virginio Jr. and Ivan Luiz Marques Ricarte

Nonadherence to Oral Antihyperglycemic Agents: Subsequent Hospitalization and Mortality Among Patients with Type 2 Diabetes in Clinical Practice 60
Vivienne J. Zhu, Wanzhu Tu, Marc B. Rosenman and J. Marc Overhage

Standards and Guidelines for Telehealth and Telemedicine

Validation of Minimum Data of Archetyped Telehealth Clinical Report for Monitoring Prenatal Care 64
Danielle Santos Alves, Valéria Cesário Times and Magdala de Araújo Novaes

Design, Implementation and Evaluation of Telehealth Solutions

One Million Electrocardiograms of Primary Care Patients: A Descriptive Analysis 69
Emmanuel Chazard, Milena Soriano Marcolino, Chloé Dumesnil, Alexandre Caron, Daniel Moore F. Palhares, Grégoire Ficheur, Barbara C. A. Marino, Maria Beatriz M. Alkmim, Régis Beuscart and Antonio Luiz Ribeiro

ePoint.telemed – An Open Web-Based Platform for Home Monitoring of Patients with Chronic Heart Failure 74
Halgeir Holthe and J. Artur Serrano

Mobile Healthcare System for Health Checkups and Telemedicine in Post-Disaster Situations 79
Min Hu, Megumi Sugimoto, Andrew Rebeiro Hargrave, Yasunobu Nohara, Michiko Moriyama, Ashir Ahmed, Shuji Shimizu and Naoki Nakashima

Assessing the Potential Use of Eye-Tracking Triangulation for Evaluating the Usability of an Online Diabetes Exercise System 84
Clara Schaarup, Gunnar Hartvigsen, Lars Bo Larsen, Zheng-Hua Tan, Eirik Årsand and Ole Kristian Hejlesen

Mobile Technology (m-Health), Apps and Sensors

Mobile App to Reduce Inactivity in Sedentary Overweight Women 89
Joseph Finkelstein, McKenzie Bedra, Xuan Li, Jeffrey Wood and Pamela Ouyang

iDECIDE: A Mobile Application for Insulin Dosing Using an Evidence Based Equation to Account for Patient Preferences 93
Buffy Lloyd, Danielle Groat, Curtiss B. Cook, David Kaufman and Adela Grando

Reconfigurable Embedded System for Electrocardiogram Acquisition 98
Marcel Seiji Kay and Fábio Iaione

Availability Communication: Requirements for an Awareness System to Support Nurses' Handling of Nurse Calls 103
Joakim Klemets and Pieter Toussaint

Optimizing Decision Support for Tailored Health Behavior Change Applications 108
Rita Kukafka, In cheol Jeong and Joseph Finkelstein

Towards an Ontology-Driven Framework to Enable Development of Personalized mHealth Solutions for Cancer Survivors' Engagement in Healthy Living 113
Sahiti Myneni, Muhammad Amith, Yimin Geng and Cui Tao

A Mobile and Intelligent Patient Diary for Chronic Disease Self-Management 118
William Van Woensel, Patrice C. Roy, Samina R. Abidi and Syed S.R. Abidi

Mobile Early Detection and Connected Intervention to Coproduce Better Care in Severe Mental Illness 123
Pauline Whelan, Matthew Machin, Shôn Lewis, Iain Buchan, Caroline Sanders, Eve Applegate, Charlotte Stockton, Sally Preston, Robert Andrew Bowen, Zhimin Ze, Chris Roberts, Linda Davies, Til Wykes, Nicholas Tarrier, Shitij Kapur and John Ainsworth

Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability? 127
Mobin Yasini and Guillaume Marchand

Ubiquitous (u-Health), Pervasive Computing

Towards Improving Hypertensive Patients Care: Pervasive Monitoring and Diagnosis Support 132
Jorge Céspedes and Cynthia Villalba

Web-Based Interventions, Web 2.0, Social Media and Networks

E-Patient Reputation in Health Forums 137
Amine Abdaoui, Jérôme Azé, Sandra Bringay and Pascal Poncelet

Web-Based Self-Management Support Interventions for Cancer Survivors: A Systematic Review and Meta-analyses 142
Ae Ran Kim and Hyeoun-Ae Park

The Shared Decision Making Frontier: A Feasibility and Usability Study for Managing Non-Critical Chronic Illness by Combining Behavioural & Decision Theory with Online Technology 148
Amina Russell, William Van Woensel and Samina Raza Abidi

Patient Portals and Personal Health Records and Systems

Experiences of Healthcare Professionals to the Introduction in Sweden of a Public eHealth Service: Patients' Online Access to Their Electronic Health Records 153
Ture Ålander and Isabella Scandurra

Characterizing Patient-Generated Clinical Data and Associated Implications for Electronic Health Records 158
Elliot G. Arsoniadis, Rabindra Tambyraja, Saif Khairat, Cyrus Jahansouz, Daren Scheppmann, Mary R. Kwaan, Gretchen Hultman and Genevieve B. Melton

Robotics and Virtual Reality Systems (eg. Surgery)

Recognizing Clinical Styles in a Dental Surgery Simulator 163
Phattanapon Rhienmora, Peter Haddawy, Siriwan Suebnukarn, Poonam Shrestha and Matthew N. Dailey

Aged Care Systems and Solutions for People with Special Needs

Web-Based Auditory Self-Training System for Adult and Elderly Users of Hearing Aids 168
Simone Virginia Vitti, Wanderléia Quinhoneiro Blasca, Daniel Sigulem and Ivan Torres Pisa

Hospital and Clinical Information Systems

Design, Implementation and Evaluation of an Architecture Based on the CDA R2 Document Repository to Provide Support to the Contingency Plan 173
Fernando Campos, Daniel Luna, Dean F. Sittig and Fernán González Bernaldo de Quirós

Archetype Based Patient Data Modeling to Support Treatment of Pituitary Adenomas 178
Claire Chalopin, Dirk Lindner, Stefan Kropf and Kerstin Denecke

Evaluating Business Value of IT in Healthcare: Three Clinical Practices from Australia and the US 183
Peter Haddad, Jonathan L. Schaffer and Nilmini Wickramasinghe

Toward User-Centered Patient Safety Event Reporting System: A Trial of Text Prediction in Clinical Data Entry 188
Lei Hua and Yang Gong

The Role of Hospital Information Systems in Universal Health Coverage Monitoring in Rwanda 193
Gustave Karara, Frank Verbeke and Marc Nyssen

Bluetooth Roaming for Sensor Network System in Clinical Environment 198
Tomohiro Kuroda, Haruo Noma, Kazuhiko Takase, Shigeto Sasaki and Tadamasa Takemura

On Building an Ontological Knowledge Base for Managing Patient Safety Events 202
Chen Liang and Yang Gong

Case Study: Applying OpenEHR Archetypes to a Clinical Data Repository in a Chinese Hospital 207
Lingtong Min, Li Wang, Xudong Lu and Huilong Duan

Quality Indicators from Laboratory and Radiology Information Systems 212
Matthieu Schuers, Mehr B. Joulakian, Nicolas Griffon, Joanne Pachéco, Carine Périgard, Eric Lepage, Ludivine Watbled, Philippe Massari and Stéfan J. Darmoni

Standards for Exchanging Health Information

Building a Semantic Interoperability Framework for Care and Research in Fibromuscular Dysplasia 217
Marie-Christine Jaulent, Ariane Assélé-Kama, Sébastien Savard, Alessandra Giavarini, Emmanuel Touzé, Xavier Jeunemaître, Adrien Ugon, Pierre-François Plouin and Laurent Toubiana

Micro- and Macrointegration Profiles for Medical Devices and Medical IT Systems 222
Raluca Pahontu, Angela Merzweiler, Gerd Schneider and Björn Bergh

eHealth Standards Development Projects

Product-Based Safety Certification for Medical Devices Embedded Software 227
José Augusto Neto, Jemerson Figueiredo Damásio, Paul Monthaler and Misael Morais

Electronic Prescription and Computerized Provider Order Entry

Patient Safety at Transitions of Care: Use of a Compulsory Electronic Reconciliation Tool in an Academic Hospital 232
Daniel A. Rizzato Lede, Sonia E. Benítez, John C. Mayan III, María I. Smith, Analía J. Baum, Daniel R. Luna and Fernán González Bernaldo de Quirós

Lead User Design: Medication Management in Electronic Medical Records 237
Morgan Price, Jens H. Weber, Iryna Davies and Paule Bellwood

High Override Rate for Opioid Drug-Allergy Interaction Alerts: Current Trends and Recommendations for Future 242
Maxim Topaz, Diane L. Seger, Kenneth Lai, Paige G. Wickner, Foster Goss, Neil Dhopeshwarkar, Frank Chang, David W. Bates and Li Zhou

Nursing Information Systems

Usability Testing of PROCEnf-USP: A Clinical Decision Support System 247
Heloisa Helena Ciqueto Peres, Diná de Almeida Lopes Monteiro da Cruz, Michelle Tellez, Rita de Cassia Gengo e Silva, Diley Cardoso Franco Ortiz and Regina Celia dos Santos Diogo

Patient Safety in Critical Care Unit: Development of a Nursing Quality Indicator System 251
Camila S.P. Lima and Sayonara F.F. Barbosa

Pre-Implementation Study of a Nursing e-Chart: How Nurses Use Their Time 255
Maria B. Schachner, Francisco J. Recondo, Janine A. Sommer, Zulma A. González, Gabriela M. García, Daniel R. Luna and Sonia E. Benítez

Clinical Decision Support and Guideline Systems and Protocols

Web-Tool to Support Medical Experts in Probabilistic Modelling Using Large Bayesian Networks With an Example of Hinosinusitis 259
Mario A. Cypko, David Hirsch, Lucas Koch, Matthaeus Stoehr, Gero Strauss and Kerstin Denecke

Physicians' Attitudes Towards the Advice of a Guideline-Based Decision Support System: A Case Study With OncoDoc2 in the Management of Breast Cancer Patients 264
Jacques Bouaud, Jean-Philippe Spano, Jean-Pierre Lefranc, Isabelle Cojean-Zelek, Brigitte Blaszka-Jaulerry, Laurent Zelek, Axel Durieux, Christophe Tournigand, Alexandra Rousseau, Pierre-Yves Vandenbussche and Brigitte Séroussi

Use of a Proven Framework for Computer Decision Support Within the Intermountain Healthcare Network 270
R. Scott Evans, James F. Lloyd, Kyle V. Johnson, Stephen Howe and Jacob S. Tripp

An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders 275
Alexandre Galopin, Jacques Bouaud, Suzanne Pereira and Brigitte Seroussi

Understanding Deviations from Clinical Practice Guidelines in Adult Soft Tissue Sarcoma 280
Esther Goldbraich, Zeev Waks, Ariel Farkash, Marco Monti, Michele Torresani, Rossella Bertulli, Paolo Giovanni Casali and Boaz Carmeli

INITIATE: An Intelligent Adaptive Alert Environment 285
Borna Jafarpour, Samina Raza Abidi, Ahmad Marwan Ahmad and Syed Sibte Raza Abidi

Influence Diagram as a Support Tool for Clinical Decisions in Cardiopulmonary and Metabolic Rehabilitation 290
Patrícia da Silva Klahr, Christian Corrêa Coronel, Caroline Cabral Robinson, João Marcelo Fonseca, Cecília Dias Flores and Rodrigo Della Méa Plentz

Advances in Infection Surveillance and Clinical Decision Support with Fuzzy Sets and Fuzzy Logic 295
Walter Koller, Jeroen S. de Bruin, Andrea Rappelsberger and Klaus-Peter Adlassnig

A Decision Fusion Framework for Treatment Recommendation Systems 300
Jing Mei, Haifeng Liu, Xiang Li, Guotong Xie and Yiqin Yu

Health Care Decision Support System for the Pediatric Emeregency Department Management 305
Sarah Ben Othman, Slim Hammadi, Alain Quilliot, Alain Martinot and Jean-Marie Renard

eHealth Tools for Health Authorities and Professionals

Analysis of Hospital Processes with Process Mining Techniques 310
Arturo Orellana García, Damián Pérez Alfonso and Osvaldo Ulises Larrea Armenteros

Improving Hospital-Wide Early Resource Allocation Through Machine Learning 315
Daniel Gartner and Rema Padman

An eHealth Approach to Reporting Allergic Reactions to Food and Closing the Knowledge Gap 320
Christopher Munro, Aida Semic-Jusufagic, Katarzyna Pyrz, Philip Couch, Audrey Dunn-Galvin, Niels Peek, Marina Themis, Clare Mills, Iain Buchan, Jonathan Hourihane and Angela Simpson

PancreApp: An Innovative Approach to Computational Individualization of Nutritional Therapy in Chronic Gastrointestinal Disorders 325
Konrad Stawiski, Alicja Strzałka, Anna Puła and Krzysztof Bijakowski

An Intelligent Ecosystem for Providing Support in Prehospital Trauma Care in Cuenca, Ecuador 329
Cristian Timbi-Sisalima, Edgar B. Rodas, Juan C. Salamea, Hernán Sacoto, Diana Monje-Ortega and Vladimir Robles-Bykbaev

Usability, Human-Computer Interaction, Natural User Interfaces

Quantifying the Activities of Self-Quantifiers: Management of Data, Time and Health 333
Manal Almalki, Fernando Martin Sanchez and Kathleen Gray

Mobile Usability Testing in Healthcare: Methodological Approaches 338
Elizabeth M. Borycki, Helen Monkman, Janessa Griffith and Andre W. Kushniruk

Making an APPropriate Care Program for Indigenous Cardiac Disease: Customization of an Existing Cardiac Rehabilitation Program 343
DanaKai Bradford, David Hansen and Mohan Karunanithi

A Serious Game for Upper Limb Stroke Rehabilitation Using Biofeedback and Mirror-Neurons Based Training 348
Diego João Cargnin, Marcos Cordeiro d'Ornellas and Ana Lúcia Cervi Prado

Usability and Safety of Software Medical Devices: Need for Multidisciplinary Expertise to Apply the IEC 62366: 2007 353
Sabrina Bras Da Costa, M.C. Beuscart-Zéphir, J-M. Christian Bastien and Sylvia Pelayo

Evidence-Based Heuristics for Evaluating Demands on eHealth Literacy and Usability in a Mobile Consumer Health Application 358
Helen Monkman, Janessa Griffith and Andre W. Kushniruk

Evaluating the Impact of Player Experience in the Design of a Serious Game for Upper Extremity Stroke Rehabilitation 363
Marcos Cordeiro d'Ornellas, Diego João Cargnin and Ana Lúcia Cervi Prado

User-Centered Design of Health Care Software Development: Towards a Cultural Change 368
Enrique Stanziola, María Quispe Uznayo, Juan Marcos Ortiz, Mariana Simón, Carlos Otero, Fernando Campos and Daniel Luna

e-Learning Environments and MOOC

Online Continuing Medical Education for the Latin American Nephrology Community 372
Alvaro Margolis, Francisco Gonzalez-Martinez, Oscar Noboa, Mario Abbud-Filho, Leticia Lorier, Marcelo Nin, Ricardo Silvariño, Sofia García, Jacqueline Pefaur, Gustavo C. Greloni, Irene L. Noronha, Antonio Lopez, María A. Ribeiro-Alves, Roberto Tanús and Juan Fernández-Cean

Citizens' Access to Health and Wellness Information

Danish Citizens and General Practitioners' Use of ICT for their Mutual Communication 376
Pernille Bertelsen and Lone Stub Petersen

The Online Availability of Multilingual Health Promotion Materials Produced by Local Health Departments: An Information Assessment 380
Daniel Capurro, Shomir Chaudhuri and Anne M. Turner

Serious Games: A Concise Overview on What They Are and Their Potential Applications to Healthcare 386
Guido Giunti, Analía Baum, Diego Giunta, Fernando Plazzotta, Sonia Benitez, Adrián Gómez, Daniel Luna and Fernan González Bernaldo de Quiros

Living with Lung Cancer – Patients' Experiences as Input to eHealth Service Design 391
Maria Hägglund, Peter Bolin and Sabine Koch

Applying a Geospatial Visualization Based on USSD Messages to Real Time Identification of Epidemiological Risk Areas in Developing Countries: A Case of Study of Paraguay 396
Silvia Ochoa, Julia Talavera and Julio Paciello

Using Publicly Available Data to Characterize Consumers Use of Email to Communicate with Healthcare Providers 401
Ryan H. Sandefer, Saif S. Khairat, David S. Pieczkiewicz and Stuart M. Speedie

Empirical Research and Evaluation

Health Information Technology Evaluation Framework (HITREF) Comprehensiveness as Assessed in Electronic Point-of-Care Documentation Systems Evaluations 406
Paulina S. Sockolow, Kathryn H. Bowles and Michelle Rogers

Economic and Cost-Effectiveness Analysis

Patient Outcomes as Transformative Mechanisms to Bring Health Information Technology Industry and Research Informatics Closer Together 410
Jacob Krive

Domain Analysis of Integrated Data to Reduce Cost Associated with Liver Disease 414
Tasneem Motiwala, Bobbie Kite, Kelly Regan, Gregg M. Gascon and Philip R.O. Payne

Socio-Organizational Impact, Quality Assessment and Improvement

The Case for Conceptual and Computable Cross-Fertilization Between Audit and Feedback and Clinical Decision Support 419
Benjamin Brown, Niels Peek and Iain Buchan

Inside the Black Box of Audit and Feedback: a Laboratory Study to Explore Determinants of Improvement Target Selection by Healthcare Professionals in Cardiac Rehabilitation 424
Wouter T. Gude, Sabine N. van der Veer, Mariëtte M. van Engen-Verheul, Nicolette F. de Keizer and Niels Peek

Using Patient Journey Modelling to Visualise the Impact of Policy Change on Patient Flow in Community Care 429
Rosemary Sayvong and Joanne Curry

Barriers to Clinical System Implementation

Challenges and Hurdles of eHealth Implementation in Developing Countries 434
H.F. Mandirola Brieux, J.H. Bhuiyan Masud, S. Kumar Meher, V. Kumar, F. Portilla, S. Indarte, D. Luna, C. Otero, P. Otero and F. González Bernaldo de Quirós

On the Development of a Hospital-Patient Web-Based Communication Tool: A Case Study From Norway 438
Conceição Granja, Kari Dyb, Stein Roald Bolle and Gunnar Hartvigsen

Barriers and Facilitators to the Introduction of Digital Pathology for Diagnostic Work 443
Rebecca Randell, Roy A. Ruddle and Darren Treanor

Informatics Policies and Ethical Issues

Big Data Clinical Research: Validity, Ethics, and Regulation 448
E. Andrew Balas, Marlo Vernon, Farah Magrabi, Lynne Thomas Gordon and Joanne Sexton

Streaming Physiological Data: General Public Perceptions of Secondary Use and Application to Research in Neonatal Intensive Care 453
Carolyn McGregor, Jennifer Heath and Yvonne Choi

Driving the Profession of Health Informatics: The Australasian College of Health Informatics 458
Christopher Pearce, Klaus Veil, Peter Williams, Andrew Cording, Siaw-Teng Liaw and Heather Grain

Other Aspects Related with the Use of Information for Health

Person-Specific Standardized Vulnerability Assessment in Health and Social Care 462
Léa A. Deleris, Pól Mac Aonghusa and Robert Shorten

Trust, Perceived Risk, Perceived Ease of Use and Perceived Usefulness as Factors Related to mHealth Technology Use 467
Rebecca Schnall, Tracy Higgins, William Brown, Alex Carballo-Dieguez and Suzanne Bakken

Change Management and Projects Planning and Implementation

Electronic Health Record System Contingency Plan Coordination: A Strategy for Continuity of Care Considering Users' Needs 472
Marcela T. Fernández, Adrián R. Gómez, Américo M. Santojanni, Alfredo H. Cancio, Daniel R. Luna and Sonia E. Benítez

Flexibility First, Then Standardize: A Strategy for Growing Inter-Departmental Systems 477
Arnvør á Torkilsheyggi

Human Factors Predicting Failure and Success in Hospital Information System Implementations in Sub-Saharan Africa 482
Frank Verbeke, Gustave Karara and Marc Nyssen

National and International Health IT Projects

Implementing a National Scottish Digital Health & Wellbeing Service at Scale: A Qualitative Study of Stakeholders' Views 487
Ruth Agbakoba, Marilyn McGee-Lennon, Matt-Mouley Bouamrane, Nicholas Watson and Frances Mair

Evaluating a Proof-of-Concept Approach of the German Health Telematics Infrastructure in the Context of Discharge Management 492
Ursula Hübner, Georg Schulte, Björn Sellemann, Matthias Quade, Thorsten Rottmann, Matthias Fenske, Nicole Egbert, Raik Kuhlisch and Otto Rienhoff

Monitoring Telemedicine Implementation in Denmark 497
Christian Nøhr, Sidsel Villumsen, Stephanie Bernth Ahrenkiel and Lars Hulbæk

Education and Training of Health Professionals and Specialists

Clinical Informatics Board Specialty Certification for Physicians: A Global View 501
Adi V. Gundlapalli, Aditya V. Gundlapalli, William W. Greaves, Denece Kesler, Peter Murray, Charles Safran and Christoph U. Lehmann

A Pilot Study of Computer-Based Simulation Training for Enhancing Family Medicine Residents' Competence in Computerized Settings 506
Aviv Shachak, Sharon Domb, Elizabeth Borycki, Nancy Fong, Alison Skyrme, Andre Kushniruk, Shmuel Reis and Amitai Ziv

Mind the Gap: The Discrepancies Between Patient Self-Reported Quality of Life and Medical Staff-Estimated Quality of Life 511
Yumiko Shimamoto, Mai Miyabe, Shuko Shikata and Eiji Aramaki

Identifying Effective Approaches for Dissemination of Clinical Evidence – Correlation Analyses on Promotional Activities and Usage of a Guideline-Driven Interactive Case Simulation Tool in a Statewide HIV-HCV-STD Clinical Education Program 515
Dongwen Wang, Xuan Hung Le and Amneris E. Luque

Building Comprehensive and Sustainable Health Informatics Institutions in Developing Countries: Moi University Experience 520
Martin C. Were, Abraham Siika, Paul O. Ayuo, Lukoye Atwoli and Fabian Esamai

Health Informatics: Developing a Masters Programme in Rwanda Based on the IMIA Educational Recommendations and the IMIA Knowledge Base 525
Graham Wright, Frank Verbeke, Marc Nyssen and Helen Betts

Information Retrieval

Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Term Coverage 529
David T. Marc, Rui Zhang, James Beattie, Laël C. Gatewood and Saif S. Khairat

Recruit – An Ontology Based Information Retrieval System for Clinical Trials Recruitment 534
Diogo F.C. Patrão, Michel Oleynik, Felipe Massicano and Ariane Morassi Sasso

Assessing the Need of Discourse-Level Analysis in Identifying Evidence of Drug-Disease Relations in Scientific Literature 539
Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, Dingcheng Li and Hongfang Liu

InfoRoute: The CISMeF Context-Specific Search Algorithm 544
Tayeb Merabti, Romain Lelong and Stefan Darmoni

Modeling and Simulation of Human Patho/Physiology/Anatomy

A Baseline Patient Model to Support Testing of Medical Cyber-Physical Systems 549
Lenardo C. Silva, Mirko Perkusich, Hyggo O. Almeida, Angelo Perkusich, Mateus A.M. Lima and Kyller C. Gorgônio

Information Processing in Clinical Research and Trials

Leveraging Electronic Tablets and a Readily Available Data Capture Platform to Assess Chronic Pain in Children: The PROBE system 554
Vibha Anand and Steven J. Spalding

Normalization of Phenotypic Data from a Clinical Data Warehouse: Case Study of Heterogeneous Blood Type Data with Surprising Results 559
James J. Cimino

Automatic Selection of Clinical Trials Based on A Semantic Web Approach 564
Marc Cuggia, Boris Campillo-Gimenez, Guillaume Bouzille, Paolo Besana, Wassim Jouini, Jean-Charles Dufour, Oussama Zekri, Isabelle Gibaud, Cyril Garde and Regis Duvauferier

Assessing the Collective Population Representativeness of Related Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES 569
Zhe He, Shuang Wang, Elhaam Borhanian and Chunhua Weng

Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers 574
George Hripcsak, Jon D. Duke, Nigam H. Shah, Christian G. Reich, Vojtech Huser, Martijn J. Schuemie, Marc A. Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R. Rijnbeek, Johan van der Lei, Nicole Pratt, G. Niklas Norén, Yu-Chuan Li, Paul E. Stang, David Madigan and Patrick B. Ryan

Identifying Repetitive Institutional Review Board Stipulations by Natural Language Processing and Network Analysis 579
Fabrício S.P. Kury and James J. Cimino

Patient-Centered Outcomes Research in Practice: The CAPriCORN Infrastructure 584
Anthony Solomonides, Satyender Goel, Denise Hynes, Jonathan C. Silverstein, Bala Hota, William Trick, Francisco Angulo, Ron Price, Eugene Sadhu, Susan Zelisko, James Fischer, Brian Furner, Andrew Hamilton, Jasmin Phua, Wendy Brown, Samuel F. Hohmann, David Meltzer, Elizabeth Tarlov, Frances M. Weaver, Helen Zhang, Thomas Concannon and Abel Kho

Part 2

Natural-Language and Speech Processing

CA^{2}JU: An Assistive Tool for Children with Cerebral Palsy 589
Flávio Arthur O. Santos, Carlos Augusto E. M. Júnior, Hendrik Teixeira Macedo, Marco T. Chella, Rosana C. do Nascimento Givigi and Luciano Barbosa

Generation of Natural-Language Textual Summaries from Longitudinal Clinical Records 594
Ayelet Goldstein and Yuval Shahar

Classification of Contextual Use of Left Ventricular Ejection Fraction Assessments 599
Youngjun Kim, Jennifer Garvin, Mary K. Goldstein and Stéphane M. Meystre

Identification of Patients with Family History of Pancreatic Cancer – Investigation of an NLP System Portability 604
Saeed Mehrabi, Anand Krishnan, Alexandra M. Roch, Heidi Schmidt, DingCheng Li, Joe Kesterson, Chris Beesley, Paul Dexter, Max Schmidt, Mathew Palakal and Hongfang Liu

Heart Failure Medications Detection and Prescription Status Classification in Clinical Narrative Documents 609
Stéphane M. Meystre, Youngjun Kim, Julia Heavirland, Jenifer Williams, Bruce E. Bray and Jennifer Garvin

Classifying the Indication for Colonoscopy Procedures: A Comparison of NLP Approaches in a Diverse National Healthcare System 614
Olga V. Patterson, Tyler B. Forbush, Sameer D. Saini, Stephanie E. Moser and Scott L. DuVall

Automatically Expanding the Synonym Set of SNOMED CT Using Wikipedia 619
Daniel R. Schlegel, Chris Crowner and Peter L. Elkin

Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network 624
Yonghui Wu, Min Jiang, Jianbo Lei and Hua Xu

Identifying Patients with Depression Using Free-Text Clinical Documents 629
Li Zhou, Amy W. Baughman, Victor J. Lei, Kenneth H. Lai, Amol S. Navathe, Frank Chang, Margarita Sordo, Maxim Topaz, Feiran Zhong, Madhavan Murrali, Shamkant Navathe and Roberto A. Rocha

Text Mining

An Approach for Automatic Classification of Radiology Reports in Spanish 634
Viviana Cotik, Darío Filippo and José Castaño

Automated Learning of Temporal Expressions 639
Douglas Redda, YiJun Shaoa, Jing Yang, Guy Divita and Qing Zeng-Treitler

Identifying Diseases, Drugs, and Symptoms in Twitter 643
Antonio Jimeno-Yepes, Andrew MacKinlay, Bo Han and Qiang Chen

Image and Biosignal Processing

Improving Preventive Healthcare with an User-Centric Mobile Tele-Monitoring Model 648
Valter Roesler, Alécio P.D. Binotto, Cirano Iochpe, Euro Bruno Palomba and Luiz A.P. Tizatto

Calculation of Cardiac Kinetic Energy Index from PET images 653
John Sims, Marco Antônio Oliveira, José Claudio Meneghetti and Marco Antônio Gutierrez

Molecular Data Processing

TCGA4U: A Web-Based Genomic Analysis Platform to Explore and Mine TCGA Genomic Data for Translational Research 658
Zhenzhen Huang, Huilong Duan and Haomin Li

Conceptual Knowledge Discovery in Databases for Drug Combinations Predictions in Malignant Melanoma 663
Kelly Regan, Satyajeet Raje, Cartik Saravanamuthu and Philip R.O. Payne

Managing OMICS-Data: Considerations for the Design of a Clinical Research IT-Infrastructure 668
Nadine Umbach, Benjamin Löhnhardt and Ulrich Sax

Data Integration

Virtual Microscopy Large Slide Automated Acquisition: Error Analysis and Validation 672
Axel V. Mancino, Alyson Roger, Domingo L. Muscolo, German L. Farfalli, Pablo D. Roitman, Luis A. Aponte-Tinao, Lucas E. Ritacco and Federico E. Milano

Do Interoperable National Information Systems Enhance Availability of Data to Assess the Effect of Scale-Up of HIV Services on Health Workforce Deployment in Resource-Limited Countries? 677
Tom Oluoch, David Muturi, Rose Kiriinya, Anthony Waruru, Kevin Lanyo, Robert Nguni, James Ojwang, Keith P. Waters and Janise Richards

Improving Clinical Decisions on T2DM Patients Integrating Clinical, Administrative and Environmental Data 682
Daniele Segagni, Lucia Sacchi, Arianna Dagliati, Valentina Tibollo, Paola Leporati, Pasqale De Cata, Luca Chiovato and Riccardo Bellazzi

Architecture to Summarize Patient-Level Data Across Borders and Countries 687
Luís A. Bastião Silva, Carlos Días, Johan van der Lei and José Luis Oliveira

Data Mining, Machine Learning, Predictive Modeling

A Cloud-Based Infrastructure for Feedback-Driven Training and Image Recognition 691
Mani Abedini, Stefan von Cavallar, Rajib Chakravorty, Matthew Davis and Rahil Garnavi

A Two-stage Dynamic Model to Enable Updating of Clinical Risk Prediction from Longitudinal Health Record Data: Illustrated with Kidney Function 696
Artur Akbarov, Richard Williams, Benjamin Brown, Mamas Mamas, Niels Peek, Iain Buchan and Matthew Sperrin

Development and Preliminary Validation of a Dynamic, Patient-Tailored Method to Detect Abnormal Laboratory Test Results 701
Paolo Fraccaro, Benjamin Brown, Mattia Prosperi, Matthew Sperrin, Iain Buchan and Niels Peek

Automated Detection of Postoperative Surgical Site Infections Using Supervised Methods with Electronic Health Record Data 706
Zhen Hu, Gyorgy J. Simon, Elliot G. Arsoniadis, Yan Wang, Mary R. Kwaan and Genevieve B. Melton

Mining Temporal and Data Constraints Associated with Outcomes for Care Pathways 711
Xiang Li, Haifeng Liu, Jing Mei, Yiqin Yu and Guotong Xie

Acquiring Plausible Predications from MEDLINE by Clustering MeSH Annotations 716
Jose Antonio Miñarro-Giménez, Markus Kreuzthaler, Johannes Bernhardt-Melischnig, Catalina Martínez-Costa and Stefan Schulz

Effects of Plasma Transfusion on Perioperative Bleeding Complications: A Machine Learning Approach 721
Che Ngufor, Dennis Murphree, Sudhindra Upadhyaya, Nageswar Madde, Daryl Kor and Jyotishman Pathak

On the Automated Segmentation of Epicardial and Mediastinal Cardiac Adipose Tissues Using Classification Algorithms 726
Érick Oliveira Rodrigues, Felipe Fernandes Cordeiro de Morais and Aura Conci

Fusing Heterogeneous Data for Alzheimer's Disease Classification 731
Parvathy Sudhir Pillai, Tze-Yun Leong and Alzheimer's Disease Neuroimaging Initiative

Towards Constructing a New Taxonomy for Psychiatry Using Self-Reported Symptoms 736
Jessica Ross, Thomas Neylan, Michael Weiner, Linda Chao, Kristin Samuelson and Ida Sim

A Multi-Relational Model for Depression Relapse in Patients with Bipolar Disorder 741
Rogerio Salvini, Rodrigo da Silva Dias, Beny Lafer and Inês Dutra

Thermal Signal Analysis for Breast Cancer Risk Verification 746
Lincoln F. Silva, Giomar O. Sequeiros, Maria Lúcia O. Santos, Cristina A.P. Fontes, Débora C. Muchaluat-Saadea and Aura Conci

Time-Series Data Analysis of Long-Term Home Blood Pressure Measurements in Relation to Lifestyle 751
Hiroshi Takeuchi, Naoki Kodama and Shingo Takahashi

The Improvement of Dental Posture Using Personalized Biofeedback 756
Bhornsawan Thanathornwong and Siriwan Suebnukarn

Using Social Connection Information to improve Opinion Mining: Identifying Negative Sentiment About HPV Vaccines on Twitter 761
Xujuan Zhou, Enrico Coiera, Guy Tsafnat, Diana Arachi, Mei-Sing Ong and Adam G. Dunn

Data Visualization

Syndromic Surveillance of Infectious Diseases Meets Molecular Epidemiology in a Workflow and Phylogeographic Application 766
Daniel Janies, Zachary Witter, Christian Gibson, Thomas Kraft, Izzet F. Senturk and Ümit Çatalyürek

Fingerprinting Biomedical Terminologies – Automatic Classification and Visualization of Biomedical Vocabularies Through UMLS Semantic Group Profiles 771
Bastien Rance, Thai Le and Olivier Bodenreider

Controlled Terminologies and Vocabularies

Harmonizing Nursing Terminologies 776
Amy Coenen, Fernanda Paese, Virginia Saba, Kay Jansen, Nicholas R. Hardiker and Tae Youn Kim

Developing a National-Level Concept Dictionary for EHR Implementations in Kenya 780
Aggrey Keny, Steven Wanyee, Daniel Kwaro, Edwin Mulwa and Martin C. Were

Evaluation of Herbal and Dietary Supplement Resource Term Coverage 785
Nivedha Manohar, Terrance J. Adam, Serguei V. Pakhomov, Genevieve B. Melton and Rui Zhang

Semantic Alignment Between ICD-11 and SNOMED CT 790
Jean-Marie Rodrigues, David Robinson, Vincenzo Della Mea, James Campbell, Alan Rector, Stefan Schulz, Hazel Brear, Bedirhan Üstün, Kent Spackman, Christopher G. Chute, Jane Millar, Harold Solbrig and Kristina Brand Persson

Ontologies, Knowledge Bases, Data Models and Metadata

Extending the Coverage of Phenotypes in SNOMED CT Through Post-Coordination 795
Ferdinand Dhombres, Rainer Winnenburg, James T. Case and Olivier Bodenreider

A Model-Driven Approach to Customize the Vocabulary of Communication Boards: Towards More Humanization of Health Care 800
Natália M. Franco, Gabriel F. Medeiros, Edson A. Silva, Angela S. Murta, Aydano P. Machado and Robson N. Fidalgo

Pain Documentation: Validation of a Reference Model 805
Emily Gesner, Sarah A. Collins and Roberto Rocha

Consumer Health Information Needs and Question Classification: Analysis of Hypertension Related Questions Asked by Consumers on a Chinese Health Website 810
Haihong Guo, Jiao Li and Tao Dai

Generating and Executing Complex Natural Language Queries Across Linked Data 815
Thierry Hamon, Fleur Mougin and Natalia Grabar

Analyzing Operative Note Structure in Development of a Section Header Resource 821
Genevieve B. Melton, Yan Wang, Elliot Arsoniadis, Serguei V.S. Pakhomov, Terrence J. Adam, Mary R. Kwaan, David A. Rothenberger and Elizabeth S. Chen

A Process for the Representation of openEHR ADL Archetypes in OWL Ontologies 827
Alex Mateus Porn, Leticia Mara Peres and Marcos Didonet Del Fabro

Harmonizing SNOMED CT with BioTopLite: An Exercise in Principled Ontology Alignment 832
Stefan Schulz and Catalina Martínez-Costa

Standardized Cardiovascular Quality Assurance Forms with Multilingual Support, UMLS Coding and Medical Concept Analyses 837
Julian Varghese, Sarah Schulze Sünninghausen and Martin Dugas

Knowledge Acquisition and Processing

vizHOME – A Context-Based Home Assessment: Preliminary Implications for Informatics 842
Gail R. Casper, Patricia Flatley Brennan, Jesse O. Perreault and Alex G. Marvin

The Use of Applications in Distance Education Specialization Course as a Support Tool for Students Living in Remote Areas Without Internet 847
Ana Emília F. Oliveira, Rômulo M. França, Eurides F. Castro Júnior, Deborah C. L. Baesse, Mariana F. L. Maia and Elza B. Ferreira

An Architecture for Continuous Data Quality Monitoring in Medical Centers 852
Gregor Endler, Peter K. Schwab, Andreas M. Wahl, Johannes Tenschert and Richard Lenz

A Conceptual Framework for Decision-Making Support in Uncertainty- and Risk-Based Diagnosis of Rare Clinical Cases by Specialist Physicians 857
Adriano A. Santos, J. Antão B. Moura and Joseana Macêdo Fechine Régis de Araújo

A Hybrid Approach Using Case-Based Reasoning and Rule-Based Reasoning to Support Cancer Diagnosis: A Pilot Study 862
Renata M. Saraiva, João Bezerra, Mirko Perkusich, Hyggo Almeida and Clauirton Siebra

Health Information Portals and Digital Libraries

Identifying Clinical Study Types from PubMed Metadata: The Active (Machine) Learning Approach 867
Adam G. Dunn, Diana Arachi and Florence T. Bourgeois

Posters

Electronic Health Records

Characterization of Help Desk Issues After the Implementation of an Emergency Department Electronic Health Record 875
Daniel Capurro, Mauricio Soto, Patricio Giacaman and Silvia Catalán

The e-NutriHS: A Web-Based System for a Brazilian Cohort Study 876
Luciana D. Folchetti, Isis T. da Silva, Bianca de Almeida Pititto and Sandra R.G. Ferreira

Problem Oriented Medical Record: Characterizing the Use of the Problem List at Hospital Italiano de Buenos Aires 877
Mariano Franco, Maria Victoria Giussi Bordoni, Carlos Otero, Mariana Clara Landoni, Sonia Benitez, Damian Borbolla and Daniel Luna

Secretaries' Role in EHR Documentation and the Implications of Establishing a Structured EHR System 878
Monika Alise Johansen, Åse-Merete Pedersen and Gunnar Ellingsen

Development of Markup Language for Medical Record Charting: A Charting Language 879
Won-Mo Jung, Younbyoung Chae and Bo-Hyoung Jang

Using Electronic Medical Record Data to Improve HIV Patient Monitoring, Clinical Decision-Making, and Quality Improvement: Lessons from Rwanda 880
Nadine Karema, Anatole Manzi, Erick Gaju, Michèle Kayiganwa, Eric Remera, Alphonse Nshimyiryo, Gad Niyibizi, Adeline Dukuze, Neil Gupta and Cheryl Amoroso

Restructuring an EHR System and the Medical Markup Language (MML) Standard to Improve Interoperability by Archetype Technology 881
Shinji Kobayashi, Naoto Kume and Hiroyuki Yoshihara

Validation for Accuracy of Cancer Diagnosis in Electronic Medical Records Using a Text Mining Method 882
Yura Lee, Soo-Yong Shin, Sung-Min Ahn, Jae-Ho Lee and Woo-Sung Kim

Developing an Electronic Medical Record for Interlinked Care Services in Haiti 883
A. Marcelin, C. Perodin, C. Baja, A. Bright, J. Duperval, M. Duplan, F. Dérilus, S. Duda and J. Pape

Applicability of Different Types of Patient Records for Patient Recruitment Systems 884
Björn Schreiweis and Björn Bergh

Evaluating the Data Completeness in the Electronic Health Record After the Implementation of an Outpatient Electronic Health Record 885
Mauricio Soto, Daniel Capurro and Silvia Catalán

Health Information Needs for Child-in-Care 886
Cori Thompson

Standards and Guidelines for Telehealth and Telemedicine

Video Conferencing Services in Healthcare: One Communication Platform to Support All 887
Andrius Budrionis, Per Hasvold, Gunnar Hartvigsen and Johan Gustav Bellika

Design, Implementation and Evaluation of Telehealth Solutions

Implementation of a Teleconsultation Service in the Primary Health Care in Brazil 888
Rodrigo V. Andreão, Marcelo Q. Schimidt, Thiago D. Sarti and Solange L. Corradi

Introducing Home Blood Pressure Telemonitoring for Children with Hypertension 889
McKenzie Bedra and Joseph Finkelstein

Providers Expectations on Telemedicine: A Qualitative Research in a Large Healthcare Network of Latin America 890
M.V. Giussi Bordoni, F. Plazzotta, J. Sommer, S. Benítez, G. García, D. Luna and F. González B. De Quirós

Future Telehealth and Telecare Reference Design Based on IoT Technologies: From Remote Monitoring to Smart Collaborative Services with Decision Support 891
Martin Gerdes, Frank Reichert, Jan Pettersen Nytun and Rune Fensli

The EU-Project United4Health: User-Centred Design and Evaluation of a Collaborative Information System for a Norwegian Telehealth Service 892
Berglind Smaradottir, Martin Gerdes, Santiago Martinez and Rune Fensli

Enhancing Tele-Collaboration Networks by Patient Participation 893
Martin Staemmler, Heiko Münch, Uwe Engelmann and Johannes Sturm

Mobile Technology (m-Health), Apps and Sensors

The Effect of Mobile App Follow-up Care on the Number of In-Person Visits Following Ambulatory Surgery: A Randomized Control Trial 894
Kathleen Armstrong, Peter Coyte and John Semple

Impact of a Mobile Health Aplication in the Nursing Care Plan Compliance of a Home Care Service in Belo Horizonte, Minas Gerais, Brazil 895
Felipe A. de Britto, Tatiana B. Martins and Gustavo A.P. Landsberg

Augmented Reality: Real-Time Information Concerning Medication Consumed by a Patient 896
Gloria Diodati, Adrián Gómez, Marcela Martínez, Daniel Luna and Fernán González Bernaldo de Quiros

Description of a Mobile-Based Electronic Informed Consent System Development 897
Min-A Hwang and In Ja Kwak

Exploring the Challenges and Opportunities of eHealth Tools for Patients with Sickle Cell Disease 898
David-Zacharie Issom, Alexandra Zosso, Frederic Ehrler, Rolf Wipfli, Christian Lovis and Sabine Koch

Development and Usability Evaluation of the Mobile Delirium Assessment App Based on Confusion Assessment Method for Intensive Care Unit (CAM-ICU) 899
Meihua Ji, Ying Wu, Polun Chang, Xin Yang, Fangyu Yang and Shuang Xu

Viability of a Bioelectrical Signal Acquisition System Energized by Cellphone with NFC 900
Marcel Seiji Kay and Fábio Iaione

Smart Glasses – A New Tool in Medicine 901
Gunnar O. Klein, Karandeep Singh and Johan von Heideken

Access Control for Mobile Assessment Systems Using ID 902
Masaharu Nakayama, Tadashi Ishii and Kazuma Morino

Using Mobile Devices to Improve the Safety of Medication Administration Processes 903
H. Navas, L. Graffi Moltrasio, F. Ares, G. Strumia, E. Dourado and M. Alvarez

Ubiquitous (u-Health), Pervasive Computing

Study of Development for RFID System to Hospital Environment 904
Seung Kwon Hong and Myung-Whun Sung

Web-Based Interventions, Web 2.0, Social Media and Networks

Developing an Online Decision Aid for Osteoarthritis 905
Glenn Salkeld, Sally Wortley, David Hunter, Hema Umapathy and Jack Dowie

Patient Portals and Personal Health Records and Systems

An Electronic Device to Record Consensual Reflex in Human Pupil 906
H.M. Pinheiro, R.M. Costa, E.N.R. Camilo and Hua Gang

A Global Analysis of Approaches to Sharing Clinical Data with Patients 907
Jennifer E. Prey, Fernanda Polubriaginof, Gilad J. Kuperman, Victoria Tiase, Sarah A. Collins and David K. Vawdrey

Is Access to eHealth Records Important for Patients? – Opinions of Healthcare Personnel 908
Isabella Scandurra and Ture Ålander

Aged Care Systems and Solutions for People with Special Needs

Rolling Medical Practice: Ambulant Mobile Medical Care for Rural Areas 909
Jonas Schwartze, Klaus-Hendrik Wolf, Sebastian Schulz, Maike Rochon, Markus Wagner, Uwe Bannenberg, Markus Drews, Thomas Fischer, Torben Hellwig, Stefan Hofmann, Petra Höft-Budde, Ralf Jäger, Stefan Lorenz, Ruth Naumann, Maik Plischke, Jörg Reytarowski, Constanze Richter, Christiana Steinbrügge, Anja Ziegenspeck, Julius von Ingelheim and Reinhold Haux

Online Communication and Chronic Obstructive Pulmonary Disease (COPD) 910
Ina Koldkjær Sølling, Per Carøe, Kurt Lindgren and Kirsten Siggaard Mathiesen

Hospital and Clinical Information Systems

Development and Practice of ISMS at a Radiotherapy Hospital by Using IHE Integration Profiles 911
Yutaka Ando, Yuki Yoshida, Masami Mukai, Yasuo Okuda, Tadashi Kamada and Shuichi Moriguchi

Optimizing Safety, Fidelity and Usability of an Intelligent Clinical Support Tool (ICST) For Acute Hospital Care: An Australian Case Study Using a Multi-Method Delphi Process 912
Mari Botti, Bernice Redley, Lemai Nguyen, Kimberley Coleman and Nilmini Wickramasinghe

An Integrated Workflow for Secondary Use of Patient Data for Clinical Research 913
Guillaume Bouzillé, Emmanuelle Sylvestre, Boris Campillo-Gimenez, Eric Renault, Thibault Ledieu, Denis Delamarre and Marc Cuggia

Screening Mammography Efficacy: A Comparison Between Screen-Film, Computed Radiography and Digital Mammography in Taiwan 914
Abdallah Ahmed Elbakkoush, Suleman Atique and I-Jen Chiang

Evaluation of a Cyber Security System for Hospital Network 915
Mohammad A. Faysel

Design and development of an EMR for Ebola Treatment Centers in Sierra Leone Using OpenMRS 916
Darius Jazayeri, Shefali Oza, Glauber Ramos, Hamish Fraser, Jonathan M. Teich, Andrew S. Kanter and Ellen Ball

The Impact of an Electronic Medical Record on Repeat Laboratory Test Ordering Across Four Australian Hospitals 917
Andrew Georgiou, Elia Vecellio, Ling Li and Johanna I. Westbrook

Implementation of Data Drive Heart Rate and Respiratory Rate parameters on a Pediatric Acute Care Unit 918
Veena Goel, Sarah Poole, Alaina Kipps, Jonathan Palma, Terry Platchek, Natalie Pageler, Christopher Longhurst and Paul Sharek

Application of Barcoding to Reduce Error of Patient Identification and to Increase Patient's Information Confidentiality of Test Tube Labelling in a Psychiatric Teaching Hospital 919
Hsiu-Chu Liua, Hsing Li, Hsin-Fei Chang, Mei-Rou Lu and Feng-Chuan Chen

Open Source Software for Patient Data Management in Critical Care 920
Jacques Massaut, Nicolas Charretk, Olivia Gayraud, Rafael Van Den Bergh, Adelin Charles and Nathalie Edema

Measuring ICTs Adoption in Health Care Facilities in Uruguay 921
Cecilia Muxí, Juan Gil, Elisa Martínez, Betania Arispe, Rossana Occhiuzzi, Juan Lozano and Jorge Forcella

Syntactic and Semantic Errors in Radiology Reports Associated with Speech Recognition Software 922
Michael D. Ringler, Brian C. Goss and Brian J. Bartholmai

Diagnostic Imaging Integrated Network: A Teleradiology Pilot in Public Hospitals in Uruguay 923
Lilian Acosta and Karime Ruibal Faral

First Step to Big Data Research in Hospital 924
Yongdon Shin, Changmin Choi, JaeHo Lee and Soo-Yong Shin

Implementing Georeferencing in the Decision-Making Process of a Health Care Provider in Uruguay 925
Elizabeth Silva Layes, Fernando Morales and Julio Leivas

A Development of Automatic Audit System for Written Informed Consent Using Machine Learning 926
Hitomi Yamada, Tadamasa Takemura, Takahiro Asai, Kazuya Okamoto, Tomohiro Kuroda and Shigeki Kuwata

Service Quality: A Main Determinant Factor for Health Information System Success in Low-Resource Settings 927
Binyam Tilahun and Fleur Fritz

Taming the Data Quality Dragon – A Theory and Method for Data Quality by Design 928
Jens H. Weber, Morgan Price and Iryna Davies

Medical Image Management, Including Standards

A Medical Image Backup Architecture Based on a NoSQL Database and Cloud Computing Services 929
Luan Henrique Santos Simões de Almeida and Marcelo Costa Oliveira

Digital Imaging and Electronic Data Capture in Multi-Center Clinical Trials 930
Thomas M. Deserno, Verena Deserno, Daniel Haak and Klaus Kabino

Standards for Exchanging Health Information

An Exponential Increase in Regional Health Information Exchange with Collaborative Policies and Technologies 931
N. Lance Downing, Steven Lane, Mathew Eisenberg, Christopher Sharp, Jonathan Palma, Northern California HIE Working Group and Christopher Longhurst

Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the Open Medical Record System 932
Suranga N. Kasthurirathne, Burke Mamlin, Grahame Grieve and Paul Biondich

Development of Clinical Database System Specialized for Heavy Particle Therapy 933
Masami Mukai, Yutaka Ando, Yuki Yokooka, Yasuo Okuda, Masayoshi Seki, Masahiro Kimura, Hiroshi Tsuji and Tadashi Kamada

Interoperable Archetypes with a Three Folded Terminology Governance 934
Rune Pederson and Gunnar Ellingsen

Standardization of Information about Birth in the Obstetric Discharge Summary 935
Zilma S. Nogueira Reis, Juliano S. Gaspar, Isaias J.R. Oliveira, Andreia C. de Souza and Thais A. Maia

Hospitalization Discharge Summary: Standardization of Information Model 936
Andréia Cristina de Souza, Claudia Moro and Zilma S. N. Reis

eHealth Standards Development Projects

The Challenge of e-Health Presence on a Petroleum Platform: Using Telemedicine to Make Operation of Pre-Salt Wells a Reality 937
R. Nunes Ferreira, T. Lopes da Rosa, C. Benevenuto de Campos Lima, G. Brito Alves de Lima, P. dos Santos Ramos, T. Dias da Silva, A. Barbieri and E. Takeo Ueda

Archetype Development Process of Electronic Health Record of Minas Gerais 938
Thais Abreu Maia, Cristiana Fernandes De Muylder and Rodrigo Mendonça Queiroga

Electronic Prescription and Computerized Provider Order Entry

Unanticipated Consequences of Hospital-Based Insulin Management Order Sets 939
Eunme Cha, Sherita Hill Golden and Joseph Finkelstein

The Impact of Implementing a New Computerized Physician Order Entry (CPOE) System on Pharmaceutical Interventions in a Tertiary Brazilian Hospital 940
Vladimir Ribeiro Pinto Pizzo, Paula Brito Gonçalves, Lívia Maria Goncalves Barbosa, Debora Cecilia Mantovani Faustino de Carvalho, Margareth Ortiz de Camargo and Carlos Onofre de Lira

Nursing Information Systems

Methodology to Establish Associations Between Data and Clinical Assessment for Computerized Nursing Process in Intensive Care Units 941
Daniela Couto Carvalho Barra, Grace Teresinha Marcon Dal Sasso and Fernanda Paese

Nursing Software for Emergency Triage (NSET) 942
H.F. Mandirola Brieux, S. Guillen, F. La Rosa, C. Moreno and S. Benitez

Nursing Clinical Documentation System Structured on NANDA-I, NOC, and NIC Classification Systems 943
Heloisa Helena C. Peres, Diná de Almeida Lopes M. da Cruz, Michelle Tellez, Rita de Cassia Gengo e Silva, Regina Celia dos S. Diogo, Diley Cardoso F. Ortiz and Dóris R. Ortiz

A Trial of Nursing Cost Accounting Using Nursing Practice Data on a Hospital Information System 944
Akiko Miyahira, Kazuko Tada, Masatoshi Ishima, Hidenori Nagao, Tadashi Miyamoto, Yoshiaki Nakagawa and Tadamasa Takemura

An Electronic Nursing Patient Care Plan Helps in Clinical Decision Support 945
C.M. Wong, S.Y. Wu, W.H. Ting, K.H. Ho, L.H. Tong and N.T. Cheung

Clinical Decision Support and Guideline Systems and Protocols

Clinical Decision Support to Implement CYP2D6 Drug-Gene Interaction 946
Pedro J. Caraballo, Mark Parkulo, David Blair, Michelle Elliott, Cloann Schultz, Joseph Sutton, Padma Rao, Jamie Bruflat, Robert Bleimeyer, John Crooks, Donald Gabrielson, Wayne Nicholson, Carolyn Rohrer Vitek, Kelly Wix, Suzette J. Bielinski, Jyotishman Pathak and Iftikhar Kullo

A Scalable Architecture for Rule Engine Based Clinical Decision Support Systems 947
Soumi Chattopadhyay, Ansuman Banerjee and Nilanjan Banerjee

Clinical Decision Support Based on Integrated Patient Models: A Vision 948
Kerstin Denecke

Impact of Specific Alerts in Potassium-Increasing Drug-Drug Interactions 949
Emmanuel Eschmann, Patrick E. Beeler and Jürg Blaser

Detecting, Monitoring, and Reporting Possible Adverse Drug Events Using an Arden-Syntax-Based Rule Engine 950
Karsten Fehre, Manuela Plössnig, Jochen Schuler, Christina Hofer-Dückelmann, Andrea Rappelsberger and Klaus-Peter Adlassnig

Systems Medicine for Multiple Myeloma: A Review on Decision Support Systems 951
Matthias Ganzinger, Christian Haux, Christian Karmen, Thomas Wetter and Petra Knaup

Bridging the Gap Between Clinical Practice Guidelines and Archetype-Based Electronic Health Records: A Novel Model Proposal 952
Diego Garcia, Claudia Maria C. Moro and Lilian Mie M. Cintho

Virtual Oncological Networks – IT Support for an Evidence-Based, Oncological Health Care Management 953
Katja Heiden, Monika Sinha and Britta Böckmann

Computer-Interpretable Clinical Guidelines: A Review and Analysis of Evaluation Criteria for Authoring Methods 954
Soudabeh Khodambashi, Laura Slaughter and Øystein Nytrø

Internal Domain-Specific Language Based on Arden Syntax and FHIR 955
Eizen Kimura and Ken Ishihara

Computationally Comparing and Analyzing All Published Scoring Systems for Diagnosis of Disseminated Intravascular Coagulation 956
Fabrício S. P. Kury and James J. Cimino

Precedent Approach to Decision Making in Clinical Processes 957
Vladimir L. Malykh and Yadulla I. Guliev

Semantic Interoperability in Clinical Decision Support Systems: A Systematic Review 958
Luis Marco-Ruiz and Johan Gustav Bellika

Development of a Mobile System Decision-Support for Medical Diagnosis of Asthma in Primary Healthcare – InteliMED 959
Júlio Menezes Jr. and Cristine Gusmão

Using Discrete Event Simulation to predict KPI's at a Projected Emergency Room 960
Pablo Concha, Liliana Neriz, Danilo Parada and Francisco Ramis

Clinical Decision Support Using Electronic Medical Records: For the Improvement of Diabetes Care and Proper Use of Insulin for Inpatients 961
Ryoma Seto and Susumu Wakabayashi

eHealth Tools for Health Authorities and Professionals

Why should I? – Acceptance of Health Information Technology Among Health Professionals 962
Joseph M.J. D'Souza and Inga Hunter

Diffusion of Innovation: Telehealth for Care at Home 963
Sharon Levy

Usability, Human-Computer Interaction, Natural User Interfaces

Visualizing Sensor Data Through an Open Platform for Connected Devices 964
Christian Bock, Thai Le, Arjmand Samuel, Danny Huang, Hilaire J. Thompson and George Demiris

Automated Evaluation of Medical Software Usage: Algorithm and Statistical Analyses 965
Ming Cao, Yong Chen, Min Zhu and Jiajie Zhang

Study of Screen Design Principles for Visualizing Medical Records 966
Kenichiro Fujita, Tadamasa Takemura and Tomohiro Kuroda

User-Centered Design to Develop Clinical Applications. Literature Review 967
Daniel Luna, María Quispe, Zulma Gonzalez, Alfredo Alemrares, Marcelo Risk, Mauro Garcia Aurelio and Carlos Otero

A Model for Usability Evaluation for the Development and Implementation of Consumer eHealth Interventions 968
David Parry, Philip Carter, Jane Koziol-McLain and Jacqueline Feather

Usability Analysis of A Customized Documentation System for Nurse Population-Health Managers 969
Lincoln Sheets, Zhijian Luan and Tina Dillon

e-Learning Environments and MOOC

Digital Inclusion for Older Adults Based on Physical Activities: An Age Concern 970
Cristine Gusmão, Júlio Menezes Jr., Carmelo Pina, Juliana Lima and João Barbosa Neto

Citizens' Access to Health and Wellness Information

Development of the eHealth Literacy Assessment Toolkit, eHLA 971
Dorthe Furstrand and Lars Kayser

Type 1 Diabetes in Twitter: Who All Listen To? 972
Elia Gabarron, Alexandra Makhlysheva and Luis Marco

Characterizing Consumer Health Informatics in Low and Middle Income Countries 973
Tugba Kutun, Miriam Föller-Nord and Thomas Wetter

3D CPR Game Can Improve CPR Skill Retention 974
Jia Li, Yimin Xu, Yahong Xu, Peng Yue, Liu Sun, Ming Guo, Shuqin Xiao, Shu Ding, Yanyan Cui, Shulan Li, Qiuying Yang, Polun Chang and Ying Wu

Improvement of Hemoglobin with Repeated Health Checks Among Women in Bangladesh 975
Mariko Nishikitani, Yasunobu Nohara, Partha Ghosh, Rafiqul Islam Maruf, Ashir Ahmed and Naoki Nakashima

Cardiac Auscultation Simulator Embedded in Virtual Learning Environment to Support Medical Teaching 976
Isaias Oliveira, Zilma Silveira N. Reis, Marilene M. Araújo and Claudia Maria V. Freire

Global Challenges in People-Centered E-Health 977
Yuri Quintana and Charles Safran

Use of Patient Portals: Personal Health Information Management in Older Adults 978
Anne Turner, Katie Osterhage, Jonathan Joe, Andrea Hartzler, Lorelei Lin and George Demiris

Machine Assisted Translation of Health Materials to Chinese: An Initial Evaluation 979
Anne M. Turner, Loma Desai, Kristin Dew, Nathalie Martin and Katrin Kirchhoff

Empirical Research and Evaluation

Lessons Learnt from Evaluation of Computer Interpretable Clinical Guidelines Tools and Methods: Literature Review 980
Soudabeh Khodambashi and Øystein Nytrø

Economic and Cost-Effectiveness Analysis

The Need for Cost-Benefit Analyses of eHealth in Low and Middle-Income Countries 981
Fleur Fritz, Mihiretu Kebede and Binyam Tilahun

Effects of Electronic Prescription on Pharmacy Productivity 982
Reima Suomi, Markus Lähteenoja and Sirpa Peura

Socio-Organizational Impact, Quality Assessment and Improvement

(Br-SCMM) Brazilian Smart City Maturity Model: A Perspective from the Health Domain 983
Ricardo Alexandre Afonso, Kellyton dos Santos Brito, Clóvis Holanda do Nascimento, Luciana Campos da Costa, Alexandre Álvaro and Vinicius Cardoso Garcia

Factors Influencing Consent for Electronic Data Linkage in Urban Latinos 984
Suzanne Bakken, Sunmoo Yoon and Niurka Suero-Tejeda

Communication Problems Between End-Users and Technicians Through a Help Desk in a Health Information System 985
Gabriela Garcia, Agustina Bertoia, Leonel Cameselle, Sonia Benitez, Diego Giunta, Analía Baum and Fernán Gonzalez Bernaldo de Quiros

Speech Therapy Teleconsultations of a Public Telehealth Service in a Developing Country 986
Aline Moreira Lucena, Maria Beatriz Alkmim, Vinícius Soares Garcia, Erica de Araújo Brandão Couto and Milena Soriano Marcolino

Teleconsultations to Provide Support for Primary Care Practitioners and Improve Quality of care – the Experience of a Large Scale Telehealth Service in Brazil 987
Milena Soriano Marcolino, Julia Pereira Afonso dos Santos, Daniel Santos Neves and Maria Beatriz Moreira Alkmim

Clinical Quality Control of a Large-Scale Teleconsultation Service 988
Maria Beatriz M. Alkmim, Milena Soriano Marcolino, Junia Xavier Maia, Cristiane G. Pessoa, Elaine Machado and Lidiane Sousa

Audit of Primary Care Electrocardiograms Sent as Emergency to a Telehealth Service – the Telehealth Network of Minas Gerais, Brazil 989
Milena S. Marcolino, Bárbara C. Carvalho, Aline M. Lucena, Ana Luiza N. França, Cristiane G. Pessoa, Daniel S. Neves and Maria Beatriz M. Alkmim

Referring Quality Assessment of Primary Health Care for Endocrinology in Rio Grande do Sul, Brazil 990
Gabriela Monteiro Grendene, Átila Szczecinski Rodrigues, Natan Katz and Erno Harzheim

2,000,000 Electrocardiograms by Distance: An Outstanding Achievement for Telehealth in Brazil 991
Milena Soriano Marcolino, Maria Beatriz Moreira Alkmim, Leonardo Bonisson, Renato Minelli Figueira and Antonio Luiz Ribeiro

Barriers to Clinical System Implementation

Care Professionals' Perceived Usefulness of A Rehabilitation Ehealth Service in Stroke Care 992
Nadia Davoody and Maria Hägglund

Routine Health Information Systems in South Africa – Opportunities for Improvement 993
Edward Nicol and Lyn A. Hanmer

Other Aspects Related with the Use of Information for Health

Perceived Reasons for High and Low Quality Observational HIV Research Data 994
Stephany N. Duda, Catherine C. McGowan and Cynthia S. Gadd

Telehealth in Brazil: Contemporary Tool for Access to Health 995
Thais Vieira Esteves and Sérgio Pacheco de Oliveira

Stigma and On-Line Health Information Seeking of U.S. South Asian Cancer Survivors 996
Sheba M. George and Marjorie Kagawa Singer

End-User Experiences and Expectations Regarding Data Registration and Reuse Before the Implementation of a (New) Electronic Health Record: A Case Study in Two University Hospitals 997
Erik Joukes, Nicolette de Keizer, Ameen Abu-Hanna, Martine de Bruijne and Ronald Cornet

Quality of Life Measurements in Spinal Cord Injury Patients 998
Enea Parimbelli, Gabriella Fizzotti, Caterina Pistarini, Carla Rognoni and Silvana Quaglini

A Methodology for Adapting Psychoeducational Content to Mobile Platforms 999
Stephanie Tucker, Sriram Iyengar and Amy Franklin

Change Management and Projects Planning and Implementation

Special People in Routine Health Information Systems Implementation in South Africa 1000
Lyn A. Hanmer and Edward Nicol

Emergency Department Information System Education and Training for Clinicians: Lessons Learned 1001
Marcelo Lopetegui, Bernd Oberpaur, Macarena Vivent, Cecilia Carrasco and Alejandro Mauro

The Role of the IT Department in Information System and Organizational Redesign 1002
Lone Stub Petersen

Open Data and Open Source Systems

Developing an Open-Source Bibliometric Ranking Website Using Google Scholar Citation Profiles for Researchers in the Field of Biomedical Informatics 1004
Dean F. Sittig, Allison B. McCoy, Adam Wright and Jimmy Lin

National and International Health IT projects

Proposal for a European Public Health Research Infrastructure for Sharing of Health and Medical Administrative Data (PHRIMA) 1005
Anita Burgun, Dina V. Oksen, Wolfgang Kuchinke, Hans-Ulrich Prokosch, Thomas Ganslandt, Iain Buchan, Tjeerd van Staa, James Cunningham, Marianne L. Gjerstorff, Jean-Charles Dufour, Jean-Francois Gibrat, Macha Nikolski, Pierre Verger, Anne Cambon-Thomsen, Cristina Masella, Emanuele Lettieri, Paolo Bertele, Marjut Salokannel, Rodolphe Thiebaut, Charles Persoz, Geneviève Chêne and Christian Ohmann

Towards a Tool for Malaria Supply Chain Management Improvement in Rural Ghana 1006
Lorena Carlo, Suzanne Bakken, Lena Mamykina, Richmond Kodie and Andrew S. Kanter

Health Interoperability into Practice: Results of the Development of a Consent Form in a Pilot Project in a Health District in São Paulo, Brazil 1007
Guilherme Becker Sander, Mauro Medeiros Borges, José Henrique do Prado Fay, Denis Costa, Augusto Cesar Gadelha Vieira, Maurício Bucciolli Guernelli, Moacyr Esteves Perche, Francisco Cantarutti, Izolda Machado Ribeiro, Cleusa Ramos Enck, Alexandre Forte Lombardi, Herberth Amaral, Edson Dota and Paulo Dornelles Picon

Disaster and Pandemic Preparedness

Healthcare in Disasters and the Role of RFID 1008
Samaneh Madanian, David Parry and Tony Norris

Biosurveillance and Population Health Monitoring

Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange 1009
Brian E. Dixon, P. Joseph Gibson, Karen Frederickson Comer and Marc Rosenman

Education and Training of Health Professionals and Specialists

The Collaborative Coordination of Special Interest Groups on the Telemedicine University Network (RUTE) in Brazil 1010
Thiago Delevidove de Lima Verde Brito, Roberto Silva Baptista, Paulo Roberto de Lima Lopes, Ana Estela Haddad, Luiz Ary Messina and Ivan Torres Pisa

A Novel Approach to Teach Medical Informatics 1011
Jürgen Holm, Thomas Bürkle, Rolf Gasenzer, François von Kaenel, Stephan Nüssli, Serge Bignens, Sang Il Kim and Michael Lehmann

Computer Experience of Nurses 1012
Luciana Schleder Gonçalves, Talita Cândida Castro and Soraya Fialek

Informatics Competencies in Nursing Management 1013
Rodrigo Jensen, Christiane Pereira Martins Casteli, Rika Miyahara Kobayashi and Maria Madalena Januário Leite

Health Informatics Competences for eHealth: What Can We Learn From a Bibliometric Analysis? 1014
Peter Kokol, Helena Blažun and Kaija Saranto

Education, Technology and Health Literacy 1015
Kurt Lindgren, Ina Koldkjær Sølling, Per Carøe and Kirsten Siggaard Mathiesen

Current Status for Teaching Nursing Informatics in Denmark, Canada, and Australia 1016
Inge Madsen, Elizabeth Cummings and Elizabeth M. Borycki

Trends of Patient Safety Topics Addressed in the Past Five MEDINFO Congresses 1017
Qiyan Zhang and Hiroshi Takeda

Social Network and Health Researchers and Professionals Mobility in Africa: Lessons Learned from AFRICA BUILD Project 1018
S.T. Traore, A. Anne, A. Khalifa, S. Bosomprah, F. Caroline, A.K. Cuzin-Kihl, B. Ingelbeen, M. Ramirez-Robles, M. Sangare, M. Niang and C.O. Bagayoko

Information Retrieval

Middleware Supporting Next Generation Data Analytics in Australia 1019
Douglas I. R. Boyle

Proposal of Local Automatic Weighing Attribute in CBIR 1020
David Jones Ferreira de Lucena, Marcelo Costa Oliviera and Aydano Pamponet Machado

Ontology-Driven Semantic Search for Brazilian Portuguese Clinical Notes 1022
Sadid A. Hasan, Xianshu Zhu, Joey Liu, Claudia M. Barra, Lucas Oliveira and Oladimeji Farri

Use of Self-Service Query Tools Varies by Experience and Research Knowledge 1023
Gregory W. Hruby, Jessica Ancker and Chunhua Weng

Comparision and Analysis of Top 10 Exercise Android Apps in Mainland China 1024
Yanling Wang, Liu Sun, Yahong Xu, Qian Xiao, Polun Chang and Ying Wu

Modeling and Simulation of Human Patho/Physiology/Anatomy

3D Printed Models and Navigation for Skull Base Surgery: Case Report and Virtual Validation 1025
Lucas E. Ritacco, Federico Di Lella, Axel Mancino, Fernan Gonzalez Bernaldo de Quiros, Carlos Boccio and Federico E. Milano

Accuracy of Chest Wall Tumor Resection Guided by Navigation: Experimental Model 1026
Lucas E. Ritacco, David E. Smith, Axel V. Mancino, German L. Farfalli, Luis Alberto Aponte-Tinao and Federico E. Milano

Information Processing in Clinical Research and Trials

Identification of Incidental Pulmonary Nodules in Free-text Radiology Reports: An Initial Investigation 1027
Lucas Oliveira, Ranjith Tellis, Yuechen Qian, Karen Trovato and Gabe Mankovich

Follow-Up Recommendation Detection on Radiology Reports with Incidental Pulmonary Nodules 1028
Lucas Oliveira, Ranjith Tellis, Yuechen Qian, Karen Trovato and Gabe Mankovich

Clinical Trial Feasibility Study Questionnaire Analysis 1029
Iñaki Soto-Rey, Martin Dugas and Fleur Fritz

Natural-Language and Speech Processing

Extraction of Adverse Events from Clinical Documents to Support Decision Making Using Semantic Preprocessing 1030
Jan Gaebel, Till Kolter, Felix Arlt and Kerstin Denecke

Development and Evaluation of Task-Specific NLP Framework in China 1031
Caixia Ge, Yinsheng Zhang, Zhenzhen Huang, Zheng Jia, Meizhi Ju, Huilong Duan and Haomin Li

Extracting Dependence Relations from Unstructured Medical Text 1032
Charles Jochim, Yassine Lassoued, Bogdan Sacaleanu and Léa A. Deleris

A Frequency-Based Strategy of Obtaining Sentences from Clinical Data Repository for Crowdsourcing 1033
Dingcheng Li, Majid Rastegar Mojarad, Yanpeng Li, Sunghwan Sohn, Saeed Mehrabi, Ravikumar Komandur Elayavilli, Yue Yu and Hongfang Liu

Extraction of Vital Signs from Clinical Notes 1035
Olga V. Patterson, Makoto Jones, Yiwen Yao, Benjamin Viernes, Patrick R. Alba, Theodore J. Iwashyna and Scott L. DuVall

Translating ICD-11 into French Using Lexical-Based Approach: A Preliminary Study 1036
Tayeb Merabti, Julien Grosjean, Jean-Marie Rodrigues and Stefan Jacques Darmoni

Text Mining

Text Mining and Data Modeling of Karyotypes to Aid in Drug Repurposing Efforts 1037
Zachary B. Abrams, Andrea L. Peabody, Nyla A. Heerema and Philip R. O. Payne

Rule-Based Cervical Spine Defect Classification Using Medical Narratives 1038
Yihan Deng, Mathias Jacob Groll and Kerstin Denecke

Comparing Drug-Disease Associations in Clinical Practice Guideline Recommendations and Drug Product Label Indications 1039
Tiffany I. Leung and Michel Dumontier

Automated Classification of Pathology Reports 1040
Michel Oleynik, Marcelo Finger and Diogo F. C. Patrão

Image and Biosignal Processing

Methods for Sonic Representation of ST Depression During Exercise 1041
Minodora Andor, Anca Tudor, Sorin Paralescu and George I. Mihalas

PIACS: A System for the Automatic Detection, Categorization and Comparison of Scratch-Related Skin Lesions in Dermatology 1042
Philipp Bruland, Waldemar Hänse, Fiona Schedel, Sonja Ständer and Fleur Fritz

The MFER Structure for Coding Medical Signals in Real Time 1043
Yonghee Lee, Jaehyuk Kim, Soonseok Kim, Kangwoo Lee and Dongho Kim

Adding Sound to ECG 1044
George I. Mihalas, Sorin Paralescu, Anca Tudor and Minodora Andor

An Optimized Superpixel Clustering Approach for High-Resolution Chest CT Image Segmentation 1045
Rafaelo Pinheiro da Rosa and Marcos Cordeiro d'Ornellas

Automated Image Retrieval of Chest CT Images Based on Local Grey Scale Invariant Features 1046
Marcelo Arrais Porto and Marcos Cordeiro d'Ornellas

Texture Analysis of Recurrence Plots Based on Wavelets and PSO for Laryngeal Pathologies Detection 1047
Taciana A. Souza, Vinícius J. D. Vieira, Suzete E. N. Correia, Silvana L. N. C. Costa, Washington C. de A. Costa and Micael A. Souza

Data Integration

On the Correlation Between Geo-Referenced Clinical Data and Remotely Sensed Air Pollution Maps 1048
Arianna Dagliati, Andrea Marinoni, Carlo Cerra, Paolo Gamba and Riccardo Bellazzi

Integrating Data from Multiple Sources for Data Completeness in a Web-Based Registry for Pediatric Renal Transplantation – The CERTAIN Registry 1049
Lennart Köster, Kai Krupka, Britta Höcker, Axel Rahmel, Undine Samuel, Wouter Zanen, Gerhard Opelz, Caner Süsal, Bernd Döhler, Lukasz Plotnicki, Christian D. Kohl, Petra Knaup and Burkhard Tönshoff

Development of Unified Lab Test Result Master for Multiple Facilities 1050
Naoto Kume, Kenji Suzuki, Shinji Kobayashi, Kenji Araki and Hiroyuki Yoshihara

Semantic Web Ontology and Data Integration: A Case Study in Aiding Psychiatric Drug Repurposing 1051
Chen Liang, Jingchun Sun and Cui Tao

Personalised Medicine Possible With Real-Time Integration of Genomic and Clinical Data to Inform Clinical Decision-Making 1052
Fernando Martin-Sanchez, Maureen Turner, Alice Johnstone, Leon Heffer, Naomi Rafael, Advisory Group, Tim Bakker, Natalie Thorne, Ivan Macciocca and Clara Gaff

Integration of Disease Specific Clinical and Genomics Datasets Using I2B2 Framework 1053
Mohyuddin

Bridging the Gap from Bench to Bedside – An Informatics Infrastructure for Integrating Clinical, Genomics and Environmental Data (ICGED) 1054
Mohyuddin

Designing an Innovative Data Architecture for the Los Angeles Data Resource (LADR) 1055
Sukrit Mukherjee, Robert A. Jenders and Sebastien Delta

Curating and Integrating Data from Multiple Sources to Support Healthcare Analytics 1056
Kenney Ng, Chris Kakkanatt, Michael Benigno, Clay Thompson, Margaret Jackson, Amos Cahan, Xinxin Zhu, Ping Zhang and Paul Huang

Integrated Database and Knowledge Base for Genomic Prospective Cohort Study in Tohoku Medical Megabank Toward Personalized Prevention and Medicine 1057
Soichi Ogishima, Takako Takai, Kazuro Shimokawa, Satoshi Nagaie, Hiroshi Tanaka and Jun Nakaya

Non-Integrated Information and Communication Technologies in the Kidney Transplantation Process in Brazil 1058
Alissa Peres Penteado, Rafael Fábio Maciel, João Erbs, Cristina Lucia Feijó Ortolani, Bartira Aguiar Roza and Ivan Torres Pisa

Are We Talking About the Same Patient? 1059
Khawllah Roussi, Vanessa Soussa, Karen Dunn Lopez, Abhinaya Balasubramanian, Gail M. Keenan, Michel Burton, Neil Bahroos, Barbara DiEugenio and Andrew D. Boyd

An HL7-FHIR-Based Object Model for a Home-Centered Data Warehouse for Ambient Assisted Living Environments 1060
Jonas Schwartze, Lars Jansen, Harald Schrom, Klaus-Hendrik Wolf, Reinhold Haux and Michael Marschollek

Data Curation: Improving Environmental Health Data Quality 1061
Lin Yang, Jiao Li, Li Hou and Qing Qian

Data Mining, Machine Learning, Predictive Modeling

Development of a Dynamic and Adaptive Simulator for Health 1062
Adrian L. Correa-Arango, Carolina Tamayo-Correa, David Mejía-Zapata, Edison Castrillón, Ever A. Torres-Silva, Ivan F. Luna-Gomez, Natalia Restrepo, Sebastián Vélez-Zuluga, Jose F. Florez-Arango and Jack Smith

On Analyzing Readmissions Using A Trajectory Model: Evidence From Israel 1063
Ofir Ben-Assuli, Rema Padman, Martha Bowman, Moshe Leshno and Itamar Shabtai

Automated Detection of Health Websites' HONcode Conformity: Can N-gram Tokenization Replace Stemming? 1064
Célia Boyer, Ljiljana Dolamic and Natalia Grabar

Data Science Solution to Event Prediction in Outsourced Clinical Trial Models 1065
Daniel Dalevi, Susan Lovick, Helen Mann, Paul D. Metcalfe, Stuart Spencer, Sally Hollis and David Ruau

Improving Hospital-wide Patient Scheduling Decisions by Clinical Pathway Mining 1066
Daniel Gartner, Ines V. Arnolds and Stefan Nickel

Annotation Methods to Develop and Evaluate an Expert System Based on Natural Language Processing in Electronic Medical Records 1067
Quentin Gicquel, Nastassia Tvardik, Côme Bouvry, Ivan Kergourlay, André Bittar, Frédérique Segond, Stefan Darmoni and Marie-Hélène Metzger

"Quartile" Screening Method to Analyze the Relationship Between HIS and "AEROS" in Japan 1068
Takayuki Hoshino, Nozomu Matsubara, Teruko Ueda, Yuichi Hirayama, Ayami Hoshino and Shiho Takaoka

Impact of Data Quality Assessment on Development of Clinical Predictive Models 1069
Jitendra Jonnagaddala, Siaw-Teng Liaw and Pradeep Ray

Evaluating Methods for Identifying Cancer in Free-Text Pathology Reports Using Various Machine Learning and Data Preprocessing Approaches 1070
Suranga Nath Kasthurirathne, Brian E. Dixon and Shaun J. Grannis

A Metadata Based Knowledge Discovery Methodology for Seeding Translational Research 1071
Cartik R. Kothari and Philip R.O. Payne

Restricted Versus Unrestricted Search Space: Experience from Mining a Large Japanese Database 1072
Hendrik Nienhoff, Ursula Huebner, Andreas Frey, Mareike Przysucha and Michio Kimura

Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation 1073
Yasunobu Nohara, Yoshifumi Wakata and Naoki Nakashima

A Data Mining Approach to Identify Sexuality Patterns in a Brazilian University Population 1074
Priscyla Waleska Simões, Samuel Cesconetto, Larissa Letieli Toniazzo de Abreu, Merisandra Côrtes de Mattos Garcia, José Márcio Cassettari Junior, Eros Comunello, Luciane Bisognin Ceretta and Sandra Aparecida Manenti

A Comparative Study of Bayes Net, Naive Bayes and Averaged One-Dependence Estimators for Osteoporosis Analysis 1075
Priscyla Waleska Simões, Leandro Luiz Mazzuchello, Larissa Letieli Toniazzo de Abreu, Diego Garcia, Maitê Gabriel dos Passos, Ramon Venson, Luciane Bisognin Ceretta, Ana Carolina Veiga Silva, Maria Inês da Rosa and Paulo João Martins

Trivalent Influenza Vaccine Adverse Event Analysis Based On MedDRA System Organ Classes Using VAERS Data 1076
Cui Tao, Jingcheng Du, Yi Cai and Yong Chen

Temporal Relation Extraction in Outcome Variances of Clinical Pathways 1077
Takanori Yamashita, Yoshifumi Wakata, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Brendan Franagan, Naoki Nakashima and Sachio Hirokawa

Data Visualization

Proof of Concept HTML5 Webapp: Type 2 Diabetes Risk Stratification 1078
Simon Lebech Cichosz, Mette Dencker Johansen and Ole Hejlesen

Design of a Graph-Based System for Similar Case Retrieval of Pulmonary Nodules 1079
José Raniery Ferreira Junior, Marcelo Costa Oliveira and Paulo Mazzoncini de Azevedo-Marques

Audit Trail Management System in Community Health Care Information Network 1080
Naoki Nakamura, Masaharu Nakayama, Jun Nakaya, Teiji Tominaga, Takuo Suganuma and Norio Shiratori

Utilizing Dental Electronic Health Records Data to Predict Risk for Periodontal Disease 1081
Thankam P. Thyvalikakath, Rema Padman, Karnali Vyawahare, Pratiksha Darade and Rhucha Paranjape

High-Performance and Large-Scale Computing

Distributed Parallel Computing in Data Analysis of Osteoporosis 1082
Priscyla Waleska Simões, Ramon Venson, Eros Comunello, Rogério Antônio Casagrande, Everson Bigaton, Lucas da Silva Carlessi, Maria Inês da Rosa and Paulo João Martins

Kmer-Indexer: A Fast K-Mer Indexing Program 1083
Wang Xiaolei, Qu Wubin, Zhang Chenggang and Zhao Dongsheng

Controlled Terminologies and Vocabularies

Cohort Discovery Query Optimization via Computable Controlled Vocabulary Versioning 1084
Todd A. Ferris and Tanya Podchiyska

A Statistical Analysis of Term Occurrences in Radiology Reporting 1085
Yi Hong, Jin Zhang, Ying Zhu and Xiaoying Zhou

Is it Possible to Make Everyone Talk in the Same Language? 1086
Maggie Lau, Vicky Fung, N.T. Cheung, Austen Wong, Edward Tam and Andy Wai

A Novel Approach to Create a Machine Readable Concept Model for Validating SNOMED CT Concept Post-Coordination 1087
Marcelo Lopetegui and Alejandro Mauro

Towards a Clinical Decision Support System for Drug Allergy Management: Are Existing Drug Reference Terminologies Sufficient for Identifying Substitutes and Cross-Reactants? 1088
William Ogallo and Andrew S. Kanter

Consumer Health Vocabulary: A Proposal for a Brazilian Portuguese Language 1089
Josceli Maria Tenório and Ivan Torres Pisa

Ontologies, Knowledge Bases, Data Models and Metadata

ICHI Categorial Structure: A WHO-FIC Tool for Semantic Interoperability of Procedures Classifications 1090
Syed M. Aljunid, Jean Marie Rodrigues, Linda Best, Zafar Ahmed, Hasrul Reeza Mustaffa, Béatrice Trombert, Syed M. Hamzah Aljunid, Julien Souvignet and Sukil Kim

National Governance of Archetypes in Norway 1091
Silje Ljosland Bakke

Textual Definitions in the Leukemia Domain: Methodological Guidelines for Biomedical Ontologies 1092
Amanda Damasceno de Souza and Maurício Barcellos Almeida

A Pilot Study on Modeling of Diagnostic Criteria Using OWL and SWRL 1093
Na Hong, Guoqian Jiang, Jyotishiman Pathak and Christopher G. Chute

Constructing a Graph Database for Semantic Literature-Based Discovery 1094
Dimitar Hristovski, Andrej Kastrin, Dejan Dinevski and Thomas C. Rindflesch

Development of an Ontology to Recommend Exercises from Conceptual Maps 1095
Márcia Ito and Débora Lina N. Ciriaco Pereira

Utility of Arden Syntax for Representation of Fuzzy Logic in Clinical Quality Measures 1096
Robert A. Jenders

Developing a Standards-Based Information Model for Representing Computable Diagnostic Criteria: A Feasibility Study of the NQF Quality Data Model 1097
Guoqian Jiang, Harold R. Solbrig, Jyotishman Pathak and Christopher G. Chute

A Standards-Based Semantic Metadata Repository to Support EHR-Driven Phenotype Authoring and Execution 1098
Guoqian Jiang, Harold R. Solbrig, Richard Kiefer, Luke V. Rasmussen, Huan Mo, Peter Speltz, William K. Thompson, Joshua C. Denny, Christopher G. Chute and Jyotishman Pathak

Development of an Adolescent Depression Ontology for Analyzing Social Data 1099
Hyesil Jung, Hyeoun-Ae Park, Tae-Min Song, Eunjoo Jeon, Ae Ran Kim and Joo Yun Lee

A Framework for Modeling Workflow Execution by an Interdisciplinary Healthcare Team 1100
Mounira Kezadri-Hamiaz, Daniela Rosu, Szymon Wilk, Craig Kuziemsky, Wojtek Michalowski and Marc Carrier

Evaluating a Hierarchical Clinical Event Linkage Model for Clinic-Specific Databases 1101
Justin Liu and Tran Truong

Enabling Self-Monitoring Data Exchange in Participatory Medicine 1102
Guillermo Lopez-Campos, Bahadorreza Ofoghi and Fernando Martin-Sanchez

DServO: A Peer-to-Peer-Based Approach to Biomedical Ontology Repositories 1103
Zakaria Mambone, Mahamadi Savadogo, Borlli Michel Jonas Some and Gayo Diallo

OntoMama: An Ontology Applied to Breast Cancer 1104
M.T.D. Melo, V.H.L. Gonçalves, H.D.R Costa, D.S Braga, L.B Gomide, C.S Alves and L.M. Brasil

National Healthcare Policies in Chile: An Ontological Meta-Analysis 1105
Alicia Núñez Mondaca, Arkalgud Ramaprasad and Thant Syn

Modelling the Medication Management System for Resource Limited Settings: A Formal Representation of the Prescribing and Dispensing Phases 1106
William Ogallo and Andrew S. Kanter

Real-Time Data Fusion Platforms: The Need of Multi-Dimensional Data-Driven Research in Biomedical Informatics 1107
Satyajeet Raje, Bobbie Kite, Jay Ramanathan and Philip Payne

Towards a Formal Representation of Processes and Objects Regarding the Delivery of Telehealth Services: The Telehealth Ontology (TEON) 1108
Filipe Santana, Stefan Schulz, Amadeu Campos and Magdala A. Novaes

Characterizing Health Information for Different Target Audiences 1109
Yueping Sun, Zhen Hou, Li Hou and Jiao Li

A Pilot Ontology for Healthcare Quality Indicators 1110
Pam White and Abdul Roudsari

Knowledge Acquisition and Processing

Fuzzy-Arden-Syntax-Based, Vendor-Agnostic, Scalable Clinical Decision Support and Monitoring Platform 1111
Klaus-Peter Adlassnig, Karsten Fehre and Andrea Rappelsberger

Online Training Assessment for Primary Care Professionals of the City of São Paulo 1112
Debora Cristina Alavarce and Heloisa Helena Ciqueto Peres

Allergy Risk Finder: Hypothesis Generation System for Allergy Risks via Web Service 1113
Eiji Aramaki, Shuko Shikata, Eriko Watabe, Mai Miyabe, Yasuyuki Usuda, Satsuki Ayaya and Shinichiro Kumagaya

Toward a Global eHealth Observatory for Nursing 1114
Claudia C. Bartz, Nicholas R. Hardiker and Amy Coenen

Representation of Biomedical Expertise in Ontologies: A Case Study About Knowledge Acquisition on HTLV Viruses and Their Clinical Manifestations 1115
Kátia Cardoso Coelho and Maurício Barcellos Almeida

Trigger Development for the Improvement of Neurological Patient Care 1116
Eija Kivekäs, Ulla-Mari Kinnunen, Kaisa Haatainen, Reetta Kälviäinen and Kaija Saranto

Patient Empowerment Through Personal Medical Recommendations 1117
Haridimos Kondylakis, Lefteris Koumakis, Eleni Kazantzaki, Maria Chatzimina, Maria Psaraki, Kostas Marias and Manolis Tsiknakis

A Software Tool to Analyze Clinical Workflows from Direct Observations 1118
Marco Schweitzer, Nelia Lasierra and Alexander Hoerbst

Knowledge-Based Immunosuppressive Therapy for Kidney Transplant Patients – From Theoretical Model to Clinical Integration 1119
Walter Seeling, Max Plischke, Jeroen S. de Bruin and Christian Schuh

Clinical Application of the Integrated Multicenter Discharge Summary Database 1120
Suzuki Takahiro, Doi Shunsuke, Hatakeyama Yutaka, Honda Masayuki, Matsumura Yasushi, Shimada Gen, Takasaki Mitsuhiro, Tsumoto Shusaku, Yokoi Hideto and Takabayashi Katsuhiko

Oncotherapy: A System for Requesting Chemotherapy Protocols 1121
Laura Vera Righi

Semantic Web

Linked Health Data: How Linked Data Can Help Provide Better Health Decisions 1122
Fernanda Farinelli, Maurício Barcellos de Almeida and Yóris Linhares de Souza

Health Information Portals and Digital Libraries

Facilitating Full-Text Access to Biomedical Literature Using Open Access Resources 1123
Hongyu Kang, Zhen Hou and Jiao Li

What Medical Informaticians Do With and Think About an International Medical Informatics Listserv: Member Survey Preliminary Findings 1124
Craig Kuziemsky, Martha B. Adams, Bonnie Kaplan, Kourosh Ravvaz and Ross Koppel

A Comparison Between LMS Tools to Support e-Health Educational Activities 1125
Magdala de Araújo Novaes, Gabriel Soares de Vasconcelos and Jackson Raniel Florencio da Silva

Subject Index 1127

Author Index 1135

This page intentionally left blank

Papers

This page intentionally left blank

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-3

Validating the Access to an Electronic Health Record: Classification and Content Analysis of Access Logs

Leandro Noer Alassia[a], Sonia Benítez[a], Daniel Roberto Luna[a], Fernán González Bernaldo de Quiros[a]

[a] *Health Informatics Department, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina*

Abstract

Electronic Health Records (EHRs) have made patient information widely available, allowing health professionals to provide better care. However, information confidentiality is an issue that continually needs to be taken into account. The objective of this study is to describe the implementation of rule-based access permissions to an EHR system.. The rules that were implemented were based on a qualitative study. Every time users did not meet the specified requirements, they had to justify access through a pop up window with predetermined options, including a free text option ("other justification").

A secondary analysis of a deidentified database was performed. From a total of 20,540,708 hits on the electronic medical record database, 85% of accesses to the EHR system did not require justification. Content analysis of the "Other Justification" option allowed the identification of new types of access. At the time to justify, however, users may choose the faster or less clicks option to access to EHR, associating the justification of access to the EHR as a barrier.

Keywords

Access Logs; Confidentiality; Accessibility.

Introduction

The health record serves several purposes, such as clinical documentation, transmission of information between clinicians, student instruction, knowledge generation, monitoring developments, and justifying interventions [1]. Hence, the medical record must be accessed by many individuals with very different aims.

In the age of the Electronic Health Records (EHRs), confidentiality and accessibility become relevant [2-4], particularly when users are not part of the care process. These issues impact the patient population, since patient records can contain sensitive information ranging from diseases, to data concerning sexuality and personal habits, to basic demographic information. According to estimates made by the American Health Information Management Association (AHIMA), 150 people on average, have access to a patient medical record during an inpatient episode [5]. Given this volume of access, the need to protect the confidentiality of patient information is evident and important [6].

In the United States, the Office of Civil Rights has reinforced the privacy and security of personal health information though the Health Insurance Portability and Accountability Act (HIPAA), setting standards and national regulations to protect sensitive electronic health information; thus defining access rights [7]. The Data Protection Act in the UK is the European counterpart to HIPAA [8]. HIPAA establishes two categories of acceptable access: "Treatment Payment Operations in Healthcare (TPO)", and "healthcare related". It is understood that research processes involve the generation of deidentified health records, which are excluded from HIPAA.

From a technical point of view, the Roles-Based Access Control (RBAC) model has shown to be useful [9-11]. In this model, people with a potential need for access to information are given permissions according to their credentials. In this way, information remains available if needed, but the aforementioned problem remains.

Different strategies have been employed to protect patient information without severely impacting the availability of data. The so-called "blue light button" strategy represents the concept of an alternate path, which allows the user to access information. This model is necessary in the context of health care where the unpredictable often happens. Violating the access control model established by RBAC can lead to confidentiality violation. For this reason, it should be subjected to an audit process [12]. This has been called an "optimistic security approach" [13].

Understanding that it is not possible to fully abandon the optimistic approach, we seek to generate a solution to minimize its use, not only because, once breached, the patient's confidentiality has been permanently compromised, but also because the subsequent audit process is cumbersome and costly in terms of human resources [14].

Some reports show restrictive approaches to users whose credentials do not justify access [15]. In others, contextual information is required to grant access, which follows the RBAC model [16,17], while other approaches tried to establish relationships through the use of relational algorithms and machine learning [18,19].

Our approach was to first examine how professionals perceive the ideal access model for EHRs [20]. The project was planned based on the results of that study as well as on the results of a field survey. The goal of this paper is to describe the implementation of access permissions to EHR based on rules.

Methods

Design

This is a cross-sectional study. A secondary analysis of a deidentified database was performed.

Setting

Hospital Italiano de Buenos Aires (HIBA) is a high complexity teaching hospital with 750 inpatient beds, located in Buenos Aires, Argentina. HIBA is part of a health network that includes a second hospital, 25 outpatient centers, and 50 private clinics. HIBA's workforce consists of 2,800 doctors, 2,800 healthcare related personnel, and 1,900 administrative

personnel. Since 1998 HIBA has been using an "in house" Health Information System (HIS) that includes a unique, modular, problem-oriented, and patient-centered web based electronic health record (EHR). In the EHR system, all staff with valid credentials can access and review all medical records, regardless of the nature of their duties. Each entry has a potentially traceable access log.

Figure 1 shows the steps in the project. This work analyzed different justifications and categories of EHR access.

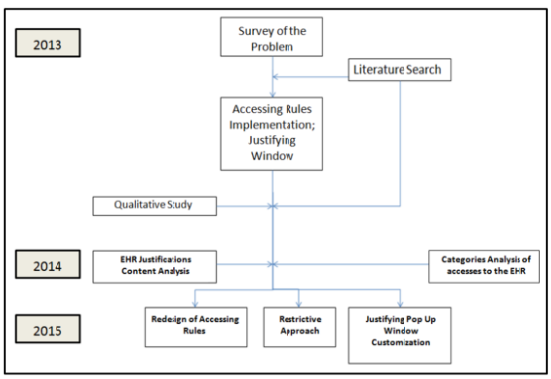


Figure 1- Data Confidentiality Project

Implementation of Access rules

Access permissions to the EHR based on specific rules were implemented in July 2013. The rules considered the most frequent use cases depending of the level of care. Ambulatory-level family physicians and physicians with a visit occurring within the last 180 days or with a scheduled visit do not need to justify access to the EHR. Attending physicians, nurses, pharmacists in the emergency department, and home care providers had the same privileges. At the inpatient level, physicians from the medical department where the inpatient was admitted, referral physicians with a referral request in the EHR, nurses working in the ward where patient is located, and pharmacists all have access to the EHR without any explanation. The qualitative study carried out in the first semester of 2013 helped to construct the different use cases [20].

Every time users did not meet the requirements listed above, they had to justify access through a pop up window with predetermined options (See Figure 2). Some options were: "medical auditor", "health informatics", "patient in my care", and "other justification". The "other justification" option allows the user to declare the reason of access in free text format. These options were chosen for different reasons. Most of them were sourced from the previous survey, in which we asked different types of users for the reasons they access the EHR. Some of them were included after the discussion, with the rest of the Medical Informatics team. We left the "other justification" option to capture not previously considered use cases. For more information, see http://www.hospitalitaliano.org.ar/infomed/index.php?contenido=ver_curso.php&id_curso=17942 .

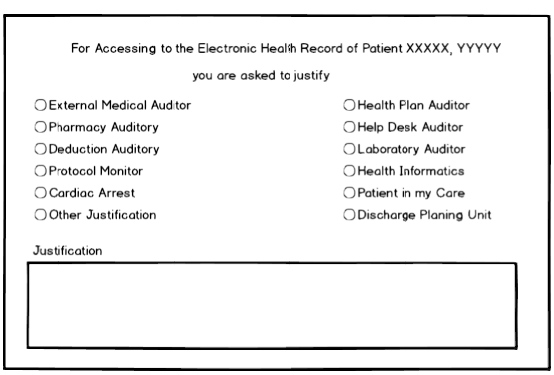


Figure 2- Pop Up Justification Window

Analysis Plan

Data from July 2013 to June 2014 were analyzed comparing both semesters, and descriptive statistics were performed. Categorical variables were presented as percentages. Free text from the "other justification" option was evaluated via content analysis. Because we did not make any adjustments in the short term, we looked for large variations occuring in the annual data over two periods.

Results

The total log accesses included 9,755,752 hits for the first period, representing an average of 62,536 hits per day, and 10,784,956 for the subsequent period (average of 59,916 daily). Of a total of 20,540,708 hits on electronic medical records, we found that 85% of access to the EHR did not need to be justified. In the remaining 15%, there were variations in the occurrences of the "patient under my care" and "other justification" options in both semesters.

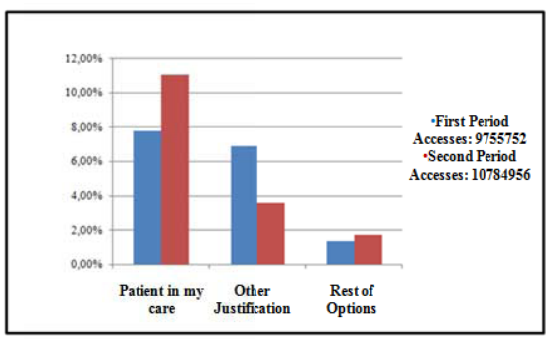


Figure 3. Remaining 15% of accesses by semester

Finally, from the content analysis of the "other justification" field, the new categories of access identified are shown in Table 1. Over all types of users, 52% of all written justifications corresponded to nonsensical text or acronyms with uncertain meaning.

Table 1 - Content analysis of the "other justification" option

Identified Case	Repetitions n=1060217 (%)
Nutrition Personnel Access	203, 693 (19.21)
Referral	131, 946 (12.44)
Billing Access	124, 681 (11.76)

Discussion

In this paper, data regarding justification of access to EHRs were analyzed. The majority (85%) of accesses responded to the rules, and users did not need to clarify the reason to access the EHR. Of the remaining 15%, the two predominant categories were "patient under my care" and "other Justification", the latter of which instructed the user to write additional information in free text.

There was a marked decrease in the selection of the "other justification" option between the first and second periods of the study. We considered that this could be due to user adaptation. This phenomenon may be considered similar to alert fatigue [21,22]. Given the repeated appearance of the pop up window, it is possible that it disrupted workflow, such that many professionals who initially justified access eventually transitioned to selecting the "patient under my care" option, which allowed entry with only two clicks. This may also explain the nonsensical text in the "other justification" option, which may have been entered in order to access records more expeditiously. Further research will help to understand the problem. On the other hand, new residents begin their training at our hospital every June/July, increasing the probability of this option being used more frequently in the first period.

With respect to the "other justification" option, content analysis allowed us to find new use cases not covered before, which led to their addition to the rules of validity for access in the next phase of the project. Comparing our experience to those in other publications, we see that the RBAC model – the major paradigm for access control over the last fifteen years – is frequently adapted to different settings. In some cases, the model is applied to make access more restrictive [15], while in others it is to allow access in the combined context of team collaboration [23], using captured information on context metadata to optimize control permissions [16,17] or even trying to predict association rules between users and patients [18]. We think that a qualitative approach followed by a content analysis of the information (provided by users, even when not formally verified because of the deidentified data use), can offer a solution to find new rules, enabling a continuous quality improvement cycle and engaging users with regard to this important topic. Further research would be useful to evaluate the accuracy of the "other justification" option, if necessary.

This study has some limitations. It is a cross sectional study, and information is from a single academic healthcare center, therefore the results cannot be generalized without validation. Additionally, we cannot ensure that the selected options in the analysis corresponded to the real reason of access, because the categories outlined in Figure 2 could be accessed by all users without any validation.

Based on data from this study, we plan to make access to the EHR more restrictive, allowing access only if the user is a family doctor or a member of a registered care team. We will maintain a "blue light button" for emergencies, accompanied by an agile and effective audit. This option will trigger a self-audit mechanism through institutional e-mail. The rules of access will be extended according to the content analysis of written justifications. Finally, we will customize the options in the pop up window according to the department and/or service where the user belongs, to allow a more accurate granularity and to reduce the need to write extra text. With these changes, we aim to improve the workflow in the newly identified use cases, improve the fidelity of structured options for justification, and reduce the need for justification in free text.

Finally, it must not to be overlooked that an EHR provides the technical infrastructure to aggregate information and establishes a longitudinal record of health information for individual patients. This accessibility of information has already opened the debate about who should grant permission for a user to access the information in the EHR. For several years, it has been postulated that the patient, as owner of the information, is the one who should define who can access their record. This permission, in turn, must be sufficiently flexible to allow the patient to set the privacy level from accessible only to a few professionals, to having no privacy restrictions, according to each owner's preferences [24]. In this sense, open questions remain about what measures can be taken to protect patients' information, such as clarifying who can access and why they are accessing the clinical data repository, while interfering as little as possible with the workflow of each healthcare professional.

Conclusion

In this analysis of access permissions to the EHR based on rules, the majority of users did not need to clarify the reason why they needed access. Rules implemented were based on a qualitative approach, and content was analyzed using information provided by users. At the time of justification, however, users may consider access justification a disruptive barrier, often choosing the faster (or "fewer clicks") option to access the information.

References

[1] Reiser, S. The clinical record in medicine. Part 1:Learning from cases. (1991)*Annals of Internal Medicine,114* (10), 902–907

[2] Bradley A Malin, Khaled El Emam, Christine M O'Keefe. Biomedical data privacy: problems, perspectives, and recent advances. J Am Med Inform Assoc. 2013 Jan-Feb; 20(1): 2–6. doi: 10.1136/amiajnl-2012-001509 PMCID: PMC3555341

[3] Gkoulalas-Divanis, A., Loukides, G., Xiong, L., & Sun, J. Informatics methods in medical privacy. (2014) Journal of Biomedical Informatics, 50, 1–3. doi:10.1016/j.jbi.2014.07.010

[4] Barrows, R. C., & Clayton, P. D. (n.d.). Privacy, confidentiality, and electronic medical records. *Journal of the American Medical Informatics Association : JAMIA*, *3*(2), 139–48.

[5] Steward, M. (2005). Electronic medical records. Privacy, confidentiality, liability. *The Journal of Legal Medicine*, *26*(4), 491–506. doi:10.1080/01947640500364762

[6] O'Brien J, Chantler C. Confidentiality and the duties of care. *Journal of Medical Ethics* 2003;29(1):36-40. doi:10.1136/jme.29.1.36

[7] U.S. Department of Health and Human Services Office for Civil Rights HIPAA Administrative Simplification Regulation Regulation Text 45 CFR Parts 160, 162, and 164 (Unofficial Version, as amended through March 26, 2013

[8] UK data protection act; 1998. http://www.legislation.gov.uk/ukpga/1998/29/ contents

[9] Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R., & Chandramouli, R. Proposed NIST standard for role-based access control. (2001).*ACM Transactions on Information and System Security*, *4*(3), 224–274. doi:10.1145/501978.501980

[10] Carrión Señor, I., Fernández Alemán, J. L., & Toval, A. Access control management in electronic health records: a systematic literature review. (2011). *Gaceta Sanitaria / S.E.S.P.A.S*, *26*(5), 463–8. doi:10.1016/j.gaceta.2011.11.019

[11] Le XH, Doll T, Barbosu M, Luque A, Wang D. Evaluation of an enhanced role- based access control model to manage information access in collaborative processes for a statewide clinical education program. J Biomed Info 2014;50: 184–95

[12] Lillian Røstad, Øystein Nytrø. Towards Dynamic Access Control for Healthcare Information Systems. Stud Health Technol Inform. 2008;136:703-8.

[13] D. Povey. Optimistic security: a new access control paradigm. Proceedings of the 1999 workshop on New security paradigms. ACM Press, Caledon Hills, Ontario, Canada, 2000. ISBN: 1581131496

[14] Boxwala, A. a, Kim, J., Grillo, J. M., & Ohno-Machado, L.Using statistical and machine learning to help institutions detect suspicious access to electronic health records. (2011) *Journal of the American Medical Informatics Association : JAMIA*, *18*(4), 498–505. doi:10.1136/amiajnl-2011-000217

[15] Lovis C, Spahni S, Cassoni N, Geissbuhler A. Comprehensive management of the access to the electronic patient record: towards trans-institutional networks Int J Med Inform. 2007 May-Jun;76(5-6):466-70. Epub 2006 Nov 3.

[16] Motta GH1, Furuie SS. A contextual role-based access control authorization model for electronic patient record. IEEE Trans Inf Technol Biomed. 2003 Sep;7(3):202-7.

[17] Koufi V, Vassilacopoulos G. Context-aware access control for pervasive access to process-based healthcare systems. Stud Health Technol Inform. 008;136:679-84.

[18] Chen K, Chang YC, Wang DW. Aspect-oriented design and implementation of adaptable access control for electronic medical records. Int J Med Inform. 2010 Mar;79(3):181-203. doi: 10.1016/j.ijmedinf.2009.12.007. Epub 2010 Feb 1.

[19] Boxwala, A. a, Kim, J., Grillo, J. M., & Ohno-Machado, L. Using statistical and machine learning to help institutions detect suspicious access to electronic health records. (2011) *Journal of the American Medical Informatics Association : JAMIA*, *18*(4), 498–505. doi:10.1136/amiajnl-2011-000217

[20] Alassia, L. N., García, M. G., Dawidowsky, A., Budalich, C. M., Benítez, S., Luna, D. R., & González Bernaldo de Quiros, F. (2014). Percepciones de los médicos frente a la legitimidad de los accesos a la historia Clínica Electrónica. In *XIV Congresso Brasileiro de Informática em Saúde. Santos, Brasil*.

[21] Jung, M., Hoerbst, a, Hackl, W. O., Kirrane, F., Borbolla, D., Jaspers, M. W. Ammenwerth, E. (2013). Attitude of physicians towards automatic alerting in computerized physician order entry systems. A comparative international survey. *Methods of Information in Medicine*, *52*(2), 99–108. doi:10.3414/ME12-02-0007

[22] Bates, D. W., Baysari, M. T., Dugas, M., Haefeli, W. E., Kushniruk, a W., Lehmann, C. U. Westbrook, J. I. (2013). Discussion of "attitude of physicians towards automatic alerting in computerized physician order entry systems". *Methods of Information in Medicine*, *52*(2), 109–27.

[23] Le, X. H., Doll, T., Barbosu, M., Luque, A., & Wang, D. (2012). An enhancement of the Role-Based Access Control model to facilitate information access management in context of team collaboration and workflow. Journal of Biomedical Informatics, 45(6), 1084–1107. http://doi.org/10.1016/j.jbi.2012.06.001

[24] Mandl, K. D., Szolovits, P., & Kohane, I. S. (2001). Public standards and patients ' control : how to keep electronic medical records accessible but private, 322(February).

Address for Correspondence

Leandro Noer Alassia, MD

Juan D. Peron 4190 (C1199ABD), Edificio DIS

Ciudad Autónoma de Buenos Aires, Argentina

leandro.alassia@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-7

A Practical Approach to Governance and Optimization of Structured Data Elements

Sarah A Collins[a,b,c], Emily Gesner[a], Steven Morgan[a], Perry Mar[a,b,c], Saverio Maviglia[a,b,c], Doreen Colburn[d], Diana Tierney[d], Roberto Rocha[a,b,c]

[a] *Clinical Informatics, Partners eCare, Partners Healthcare Systems, Boston, MA*
[b] *Department of General Internal Medicine, Brigham and Women's Hospital, Boston, MA*
[c] *Harvard Medical School, Boston, MA*
[d] *Accenture Health Practice, Accenture*

Abstract

Definition and configuration of clinical content in an enterprise-wide electronic health record (EHR) implementation is highly complex. Sharing of data definitions across applications within an EHR implementation project may be constrained by practical limitations, including time, tools, and expertise. However, maintaining rigor in an approach to data governance is important for sustainability and consistency. With this understanding, we have defined a practical approach for governance of structured data elements to optimize data definitions given limited resources. This approach includes a 10 step process: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, and 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) Monitoring for new evidence requiring revisions to reference model.

Keywords:

Structured data elements; EHR optimization; Data governance.

Introduction

A lack of consistent data capture has profound implications on reporting, e-measures, and clinical decision support and other secondary uses of EHR data [1]. Much work has been performed on the development and use of controlled terminologies to mitigate inconsistent data capture [2–8]. In an ideal EHR implementation project, the initial definition of structured data elements would require mapping to a controlled terminology *with* comparison of each newly defined data element to all other data elements that have been defined previously. Such a data governance process requires extensive resources, particularly: *training* for analysts, clinicians and informaticians in terminologies and clinical content management; *sophisticated data management tools* to continuously search, view and compare all draft and final structured data elements; and *sufficient time* to allow for iterative, analytical, and collaborative content management and validation cycles [9]. Training, tools, and time are limited resources in large scale EHR configuration and implementation projects. While best practice certainly should require tightly controlled data governance of structured data definitions at the beginning of an EHR implementation, in many cases this is not done due to limited resources, a lack of expertise, or competing priorities [9].

Challenges

A lack of consistent shared data definitions across EHR applications and clinical settings, prevent reuse and interoperability of healthcare data. EHR data collection forms defined without reference models, compromise information consistency and completeness. Adding further complexity, many healthcare organization's EHR implementation and optimization projects are quite large with limited interaction between individuals responsible for distinct applications within the EHR system [9]. These limited interactions may lead to decreased sharing of data definitions and an increase in the number of distinct data elements defined to represent similar topics.

In the absence of a pre-defined reference model for a given clinical topic, data governance requires extensive manual effort to verify completeness and consistency of documentation and this effort constantly increases with the creation of new EHR forms. This manual effort trickles down as a significant negative effect on the accuracy, timeliness, and relevance of downstream processes that use the collected data, such as billing, clinical decision support, reporting, population management, and analytics (risk estimation, prediction) [10]. Overall, a lack of well-defined data definitions that are not shared across EHR applications will require additional resources to design, build, deploy, and manage large quantities of data definitions and to remediate consequences to downstream processes [10]. Adding to the challenges, correction of data inconsistencies after data elements and forms are in active use invariably require expensive and error prone data conversions.

Potential consequences

As noted in the previous section, challenges that arise from poor data definitions are well-known, significiant and far-reaching. A solution needs to be tractable; hence, we believe focusing on measureable consequenses to help target specific efforts to improve the definitions of data across applications in an EHR is a practical approach that can be implemented in organizations with limited resources, expertise, and time.

Measureable potential consequences of poor data definitions include: regulatory and compliance requirements, billing, reporting, and clinical decision support.

Specific consequences of inconsistent data definitions that relate to regulatory and compliance requirements include:

incomplete or inadequate reporting, inconsistent evidence during auditing events, and difficulties to implement quality improvement interventions.

Billing consequences include: inconsistent evidence to confirm level of care and decreased coding accuracy and speed due to data variability. Consequences for reporting include highly complex reports that need to account for data being defined in multiple different ways, the design of reporting processes that attempt to "normalize" data to enable consistent interpretation, and the risk of not including relevant data when data elements are missed or incompatible.

Clinicial decision support (CDS) rules have to account for data being defined in multiple different ways which leads to variability and highly complex CDS design. Incomplete or incorrect data triggers can result in false positive or false negative CDS alerts and reminders. Finally, CDS maintainance requires expensive updates and re-testing to accommodate new data definitions or to remediate CDS interventions after data conversions.

Given the “real-world” resource limitations that often lead to a state of inconsistent data definitions, and the subsequent challenges and potential consequences that may result, our goal in this study was to define a practical approach for governance of structured data elements that is sustainable and repeatable. Within our proposed process, specific analytical processes (e.g., formulas, calculations, and weighted scores) require in-depth exploration of assumptions and variables and are likely dependent on an organization’s strategic aims and priorities. Therefore, detailed descriptions of analytics within our proposed process are out of scope of this paper and are planned for dissemination in a future publication.

Materials and Methods

Setting

Our method development was conducted by a workgroup comprised of informaticians, project analysts, and clinical experts. The combination of these three roles enabled a collaboration comprised of multiple expertises. For example, informaticians had expertise in data modeling, terminologies, software development lifecycle stages, and measurement research. Project analysts were highly knowledgeable about the EHR configuration and project management and helped ensure the practicality of our proposed method. Clinicians provided clinical expertise across a variety of specialties and professions to ensure the clinical relevance of our work.

Acceptance Criteria

To meet our aim of defining a practical approach for data governance for structured data elements that is sustainable and repeatable, we defined criterion for an acceptable approach. This criterion was to: 1) make efficient use of existing, though limited, resources, 2) deliver clinically relevant, reusable information, 3) reuse existing data models, 4) leverage reliable metrics for comparative decision-making, 5) manage competing priorities from various stakeholders, and 6) follow a lifecycle stages process model.

Clinical relevance was maintained by focusing on clinical topic categories as the unit of analysis. To increase reliable and practical decision-making, we defined objective measures to analyze downstream data needs and comparisons against current state. The management of competing priorities from various stakeholders was handled by prioritizing clinical topics and revisions based on consensus defined weighted scores and thresholds. Finally, each process in our practical approach was aligned with a stage from the software development lifecycle framework to provide a foundation for a repeatable approach.

Results

After iterations within our team, we identified a 10 step approach that was efficient, clinically relevant, rigorous, and practical for managing competing priorities from various stakeholders, demonstrating evidence to support decisions overtime, and ensuring sustainability for data governance needs identification. The 10 steps of our method for data gov-

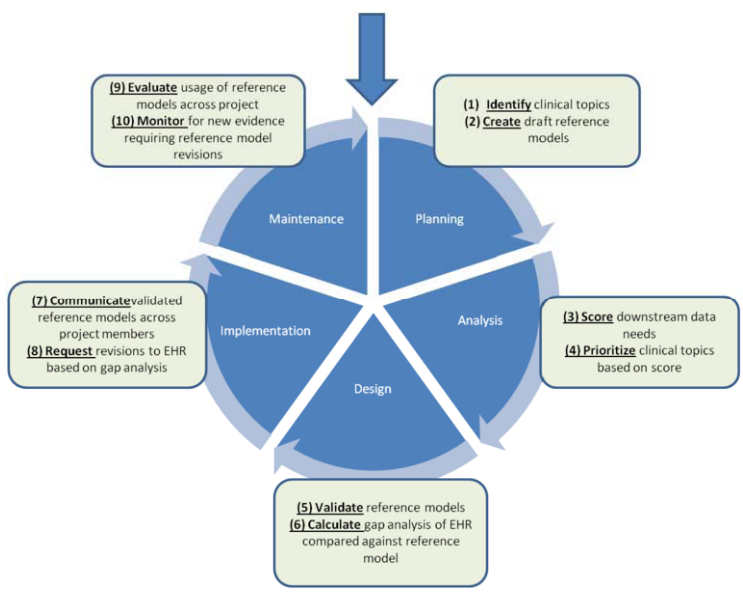


Figure 1 - Practical Approach for Data Governance for Structured Data Elements

ernance needs identification are: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, and 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) monitoring for new evidence requiring revisions to reference model.

Proposed Approach

1) Identification of clinical topics

The aim of identifying clinical topics is to align analytic focus with the strategic aims of an organization and data governance optimization. A key factor in identification of clinical topics is to select topics that are orthogonal to topics previously used to define the clinical content that is being optimized. For efficiency and minimal engagement of resources, certain assumptions may be made that topics focused on previously are a lower priority than topics that have yet to be analyzed. For example, the documentation of intravenous line placement and *urinary catheter management* may have been defined previously, but independent of each other. A focus on *catheter* placement and intravenous line *management* would allow an efficient evalution of discordant and overlapping data definitions among the topics.

2) Creation of draft reference models for clinical topics

The aim of creating a *draft* reference model is to quickly produce an artifact that can be used for initial analyses in the next stage. A reference data model for a clinical topic provides the names of each structured data element that pertains to that topic, data definitions, data types, values sets, abbreviations, synonyms, unique identifiers and original resources. We propose a targeted environmental scan of resources which publish clinical data element reference models such as: 1) HL7 Domain Analysis Models, 2) LOINC, 3) Intermountain Healthcare Clinical Element Model Browser (http://www.clinicalelement.com), 4) salient professional societies, and 5) peer-reviewed literature. This list is not intended to be exhaustive, but rather, is intended to allow for a repeatable and efficient search to identify existing reference models from organizations that are most likely to have produced them.

Existing reference models that result from the search should be compared by mapping synonymous data elements and value sets and forming a comprehensive list of data elements relevant to the clinical topic. This analysis should result in a draft reference model. In the case of retrieval of one extensively validated reference model, that should be used in its entirety.

3) Scoring of downstream data needs for clinical topics

The aim of scoring downstream data needs for clinical topics is to differentiate which clinical topics are most important for an in-depth analysis and development of a validated reference model. Given unlimited resources and time, all identified clinical topics (and in fact all content) within an EHR could be defined consistently. However, we are focused on real-world EHR implementation and optimization projects with competing priorities and limited resources; hence, a requirement of our approach is to focus and prioritize governance efforts for structured data elements.

Consistent data definitions are critical for interoperability, however, there is a lack of measureable benchmarks for interoperability for discrete clinical topics to differentiate which topics are most important to focus on. Yet, there are a set of measurable rationales for increasing the consistency of data definitions: enabling automated reporting, demonstrating adherence to regulatory requirements, maintaining clinical decision support, documenting accurate billing codes, adhering to best clinical practices and terminology described in protocols, guidelines and evidence-based literature and order sets and plans of care.

We propose an approach that identifies measurable downstream data dependencies, calculates a weighted score based on importance of each dependency and results in a total weighted score that can be used to rank the priority of each clinical topic. See Table 1 for an extract of a sample "scorecard" for this downstream data needs analysis.

4) Prioritization of clinical topics

The aim of prioritizing clinical topics is to target the highest priority topics and plan the sequence in which topics will be

Table 1 - Extract of a sample "scorecard" for downstream data needs analysis

[Clinical Topic Name] Scorecard= [total weighted Data Need Score score]

		Data Need Tallies					
Reason Data is Needed	**High Priority Data Needs**	**Data Element 1**	**Data Element 2**	**Data Element 3**	**Data Element 4**	**Data Element 5**	**Data Element 6**
Regulatory	Regulatory requirement A	✓	✓	✓	✓	✓	✓
	Regulatory requirement B	✓	✓				✓
Billing	Billing needs identified by subject matter expert	✓	✓	✓			✓
Reports	Report A		✓				
	Report B		✓				
CDS	CDS Intervention A		✓				
	CDS Intervention B			✓			
Weighted Data Need Score		[score]	[score]	[score]	[score]	[score]	[score]

addressed for data governance optimization overtime. A critical assumption is that this list is dynamic and is constantly being refined, based on scores of downstream data needs for new or re-evaluated clinical topics.

5) Validation of reference models for clinical topics

The aim of validating a reference model for a clinical topic is to produce a standard data model that is clinically accurate and complete and can be disseminated broadly for implementation. The validation of a reference data model may be completed by asking clinical experts to review the draft reference model for missing, superflous, and inaccurate information and asking analysts to review the draft reference model for any feasibility issues or technical constraints to implementation in the EHR system.

Clinical validation of reference models should be based on best practices, is consensus driven, and can be a time intensive activity. Due to the resources required validation of a reference model is done for prioritized clinical topics, and not all clinical topics.

6) Gap analysis of EHR compared against reference model

A validated reference model is useful when used as a reference standard to assert how similar or different EHR data elements are from the defined ideal state (the reference model). We have applied validated metrics for information extraction to this gap analysis [11]. These metrics can be used for a reliable gap analysis measure to identify structured documentation forms within the EHR that do not match the validated reference model. Please see Table 2 for the list of these metrics and their definitions. Documentation forms developed prior to the development of a reference model may not match exactly, but for practical purposes may be "good enough" and not a priority for revisions in the short-term. A pre-defined threshold for a total mismatch score is useful to identify the cut-off score for revisions based on mis-matched data elements. The particular threshold for this cut-off score will dependent on available resources to revise content and push changes into production for the EHR system.

7) Transparent communication of validated reference models across project members

Transparent dissemination and open and ongoing feedback about validated reference models is aimed at increasing the use of the organization's validated reference models across EHR applications and continuing efforts to optimize the models.In our proposed approach, validated reference models are actively used to conduct gap analyses. However, to make efficient use of resources all validated reference models should be communicated across all project members and available for use. Expectations should include that any newly developed content or current revisions will utilize any related validated reference models. Desired exceptions should be communicated for two purposes: 1) to clarify if information is missing or wrong in the reference model that should be added or corrected, and 2) to discover if the exception is appropriate.

8) Request EHR revisions based on gap analysis

The gap analysis provides objective and reliable measure of prioritized areas for revisions in the EHR to match validated reference models. This step to request revisions is included for consistency with change request processes that are common in many organizations. The formality of this step allows for communication about dependencies related to the requested revisions and resources needed for the revision and to manage dependencies of the revision. Additionally, this step offers an opportunity to discuss any contextual information that may impact the appropriateness of the change. For example, a gap analysis may identify 5 documentation forms that vary signficantly from Reference Model A and revisions are requested for all 5 forms. It may be identified at this stage that the 5th documentation form includes structured data elements from a research study that should not be changed during the course of the 12 month study period. A decision could then be make to make revisions for forms 1-4 now and to revise form 5 in 12 months.

9) Evaluation of usage of reference models across project

Use of an electronic collaborative tool to disseminate validated reference models would allow project members to search all validated reference models. An electronic collaboration space could enable the capture of usage metrics with little overhead by providing users the opportunity to document where and how the validated reference model was applied to the EHR application. Usage metrics of reference models can inform: 1) an understanding of the nature of clinical topics that have highly utilized reference models as a method to identify similar important topics, 2) EHR applications that are underutilizing validated reference models and may require training, 3) EHR applications that are proficient in utilizing reference models and could assist teams that are less proficient.

10) Monitoring for new evidence requiring revisions to reference model

An annual clinical subject matter expert review of validated reference models is ideal, and may be extended as limited by resources. As protocols and guidelines change, reference models with related data elements should be reviewed by a clinical subject matter expert to determine if a change is required. Consistent with our goals of a practical, efficient, and reliable approach, the downstream data analysis performed in step 3 should be re-purposed, and not replicated for this step. For example, if a pain documentation ICU protocol was updated to include new evidence, the downstream data analysis that referenced to that pain documentation ICU protocol could be easily identified and the related reference model revised to reflect new evidence.

Table 2 - Proposed Metrics for gap analysis of EHR compared against reference model

Metric	Definition
Match	Data element in the EHR is the same as the data element in Reference model
Partial Match	Data element in the EHR is different but intended to capture the same type of data as the data element in Reference model
Conflicting	Data element in EHR is not equal to the data element from Reference model
Extra	Data element is in the EHR but does not exist in Reference model
Missing	Data element is not in EHR but does exist in Reference model

Discussion

We proposed a 10 step practical approach to governance and optimization of structured data elements. Ten steps may appear to be lengthy, however, we included explicit communication and documentation steps as they are critical to effective governance. It is important to note that organizations may identify shorter paths to prioritizing clinical topics for analyses and identifying gaps in current state for redesign, however, we believe that our proposed approach most effectively combines scientific rigor, collaborative decision-making, and sustainability with efficiency.

Defining data consistently at the outset of an EHR implementation project is the ideal and our research team promotes this pro-active approach. However, we sought to provide organizations who did not or could not effecively govern consistent data definitions initially when configuring and implementing their EHR with a practical approach to remediate inconsistencies and decrease potential downstream consequences. This practical approach is being applied in our organization for a number of clinical topics and is expected to continue as an ongoing activity throughout optimization of our EHR system.

Limitations

The proposed approach is based on the experience of one integrated health system in the Northeast United States during configuration of a vendor based EHR. It is reasonable that the application of this proposed approach to governance of structured data elements would require modifications to fit within the existing infrastructure of other organizations. Our description of the proposed approach is purposefully general to increase the ability for organizations to apply the steps and concepts to meet their needs. As with the design of most best practice and governance approaches, we expect this approach to be iteratively revised and modified to fit different types of health care organizations with varied available resources and expertise.

Conclusion

We defined a 10 step practical and rigorous approach to structured data element governance in EHRs: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) monitoring for new evidence requiring revisions to reference model. This approach should be evaluated for its applicability across difference types of health care organizations and stages of EHR implementation. We believe that practical and applied approaches that streamline analytical steps while maintaining scientific rigor are highly needed and beneficial to the clinical informatics and health information technology communities amidst the great changes and adoptions that are occurring every day in health care systems throughout the world.

Acknowledgments

We would like to acknowledge all members of the Structured Data Element Workgroup at Partners eCare and all members of Partners eCare project that contributed to this work.

References

[1] (IOM) Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC: The National Academies Press; 2013.

[2] Hwang J, Cimino J, and Bakken S. Integrating Nursing Diagnostic Concepts into the Medical Entities Dictionary Using the ISO Reference Terminology Model for Nursing Diagnosis. J Am Med Informatics Assoc 2003: 10: 382–9.

[3] Kawamoto K, Del Fiol G, Strasberg HR, Hulse N, Curtis C, Cimino JJ, et al. Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard. AMIA Annu Symp Proc, 2010; pp. 377–81.

[4] Campbell JR, Xu J, and Fung KW. Can SNOMED CT fulfill the vision of a compositional terminology? Analyzing the use case for problem list. AMIA Annu Symp Proc, 2011; pp. 181–8.

[5] Chen ES, Maloney FL, Shilmayster E, and Goldberg HS. Laying the groundwork for enterprise-wide medical language processing services: architecture and process. AMIA Annu Symp Proc, 2009; pp. 97–101.

[6] Kim H, Dykes P, Mar P, Goldsmith D, Choi J, and Goldberg H. Towards a standardized representation to support data reuse: representing the ICNP semantics using the HL7 RIM. Stud Health Technol Inform 2009: 146: 308–13.

[7] Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc n.d.: 16: 81–8.

[8] Fung KW, McDonald C, and Srinivasan S. The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions. J Am Med Inf Assoc 2010: 17: 675–80.

[9] Collins S, Bavuso K, Zuccotti G, and Rocha RA. Lessons Learned for Collaborative Clinical Content Development. Appl Clin Inf 2013: 4.

[10] Cullen T, Bouhaddou O, and Goyal P. Role of Health Informatics in Success of Health IT Projects. HIMSS 2014 Annu Conf Exhib, 2014.

[11] Chinchor N, and Sundheim B. MUC-5 Evalution Metrics. Fifth Messag Underst Conf, 1992; pp. 69–78.

Address for correspondence

Sarah Collins, RN, PhD, Partners Healthcare Systems, Floor 2, 93 Worcester St, Wellesley, MA, 02445, USA

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-12

Systems Architecture for a Nationwide Healthcare System

Jorge Abin[a,b], Horacio Nemeth[a], Ignacio Friedmann[a]

[a] Agency for the e-Gov and the Information Society & Knowledge of Uruguay (AGESIC)
[b] Systems Integration Laboratory (LINS), Institute of Computer Studies (InCo), School of Engineering, University of the Republic (UdelaR)

Abstract

From a national level to give Internet technology support, the Nationwide Integrated Healthcare System in Uruguay requires a model of Information Systems Architecture. This system has multiple healthcare providers (public and private), and a strong component of supplementary services. Thus, the data processing system should have an architecture that considers this fact, while integrating the central services provided by the Ministry of Public Health. The national electronic health record, as well as other related data processing systems, should be based on this architecture. The architecture model described here conceptualizes a federated framework of electronic health record systems, according to the IHE affinity model, HL7 standards, local standards on interoperability and security, as well as technical advice provided by AGESIC. It is the outcome of the research done by AGESIC and Systems Integration Laboratory (LINS) on the development and use of the e-Government Platform since 2008, as well as the research done by the team Salud.uy since 2013.

Keywords:

eHealth Information Systems; eHealth; Interoperability; e-Government; HL7; IHE; System Architecture for Nationwide ehealth; Nationwide Healthcare Systems Architecture.

Introduction

From the perspective of conception and systems architecture, this document includes two relevant aspects that constitute the conceptual model of a technological platform to support the Electronic Health Records System at a National level (HCEN in Spanish).

As it will be stated below, this “eHealth Platform” (PS in Spanish) is implemented as an extension of the "e-Government Platform” (PGE in Spanish), reusing and expanding security components and interoperability.

In Uruguay, the development of the Electronic Health Record (EHR) at a national level is included as a goal in the Digital Agenda 2011-2015 [1]. In fact, objective number 14 includes four specific goals; two of them are relevant to this document: 1) "During 2012, to create a health data network..." 2) "Since 2012, to create and manage a platform on electronic health records..." The concept and design of the whole system architecture is based on the context where it will be applied and, obviously, it must consider the legal and technical framework. Under the heading "Context: The Health Ecosystem in Uruguay," there is an outline of the framework within which HCEN will be developed in Uruguay. The heading "Health Platform: Architecture of a Nationwide Solution" includes a description of this platform. Finally, respective conclusions and current stage of evolution will be presented.

Justification

With focus on the particular features of the Nationwide Integrated Healthcare System in Uruguay, this article shows the need of having a national system of interoperable EHR, the possibility of applying acquired knowledge and lessons learned when conceiving, designing and operating the e-Government Platform in Uruguay, the experience on health information systems, and the inherent technological challenges.

Context: The Health Ecosystem in Uruguay

Legal Framework

Act 18211, from December 13, 2007, Nationwide Integrated Healthcare System. [2]

Act 18331, from August 11, 2008, Act on the Protection of Personal Data.

Act 18335 from August 15, 2008, on rights and duties of patients and users of healthcare services. [3]

Act 18600 from September 21, 2009, which acknowledges legal effects and validity of e-documents and e-signature.

Nationwide Integrated Healthcare System (SNIS)

In Uruguay, the Nationwide Integrated Healthcare System (SNIS in Spanish) has the following features: gives universal coverage, is tax-financed, and services are delivered by networked healthcare providers. There is one public provider and multiple private providers that may give full coverage or partial coverage, depending on the services they deliver.

The underlying principle is that the users form the core and backbone of this healthcare system. The user is entitled to choose freely a healthcare institution; once the user selects a healthcare institution, the relationship user-institution shall last at least 3 years, and the SNIS is aware of this affiliation. HCEN is the main part of the Health Information System, and it is a key component of the SNIS. National authorities have imposed an additional investment levy to service providers; most of them have submitted projects on ICTs, specifically on the development of EHR.

Apart from comprehensive healthcare providers, there are partial providers, such as private medical labs, imaging clinics, non-hospital emergency services, private physician offices, or office networks among others that are authorized and operate under regulations issued by the Ministry of Public Health (MSP in Spanish), but are not part of the SNIS. Those service providers also produce relevant medical information which must be considered by HCEN. Usually, people are affiliated with comprehensive healthcare providers and mobile emergency services, and also undergo medical examinations in private offices. Pharmacy networks must also be considered. The SNIS has been operating since 2008.

Salud.uy Program

This program results from the agreement entered into the Presidency of the Republic, the Ministry of Public Health, the Ministry of Economy and Finance and AGESIC. It is aimed at the implementation of goals included in the Digital Agenda Uruguay 2011-2015.

It is focused on the development of the HCEN and the creation of the National Bank of Health Records, which will act as the institution in charge of the HCEN management after the completion of this program.

AGESIC is in charge of the implementation of Salud.uy. There is a program board formed by one representative of each agency that is part to the above mentioned agreement and the project manager. It is supported by international funding (IDB) with domestic counterparts.

Current Situation of Healthcare Providers from the Perspective of eHealth

Every healthcare provider maintains and is responsible for providing custody for health records of their affiliates.

All healthcare providers were surveyed to assess the current situation from the perspective of ICT. They were also questioned on the acceptance of EHR and its current level of incorporation. This survey revealed a great disparity in ICT maturity levels. Regarding the incorporation of EHR, most of the comprehensive providers are carrying out projects in different stages. Practically, interoperability between healthcare providers does not exist.

Problems on Interoperability, Lessons Learned and Solutions Adopted

Uruguay, through AGESIC, has widely defined the standards on interoperability, people and addresses metadata, regulations, and recommendations related to best practices on the subject [4]. Also, Uruguay has an Object Identifier (OID) managing body (UnaOid) [5].

The interoperability use made by State agencies evolved positively with the adoption of technological solutions that were not dependent on the technological maturity of agencies. Early stages, based on training courses, manuals, workshops, were complex; the level of acceptance and adoption was very slow. The PGE-connector [6] is an abstraction that encapsulates problems related to PGE connection (security and interoperability), considerably simplifying programming related to connections or service supplying to PGE. Since the inclusion of this tool, levels of acceptance and adoption increased exponentially.

Lesson learned: interoperability, in the area of health, is even more complex than e-Government transactions; standards and profiles (IHE and HL7) establish complex mechanisms that must be solved with a high level of abstraction to facilitate their early adoption.

Connectivity, Health Network

Uruguay has one data communication network, “red.uy,” which provides the connectivity infrastructure required to link State bodies in a secure way, with proper levels of service, IT security and high availability.

The health network is consistent with the model of complementation of services of the SNIS. The core is formed by red.uy and it is extended through connections with local existing health networks (public health network, networks from private providers, and public internet access for low scale users and providers). Figure 1 shows that network. Thus, a national network is set up.

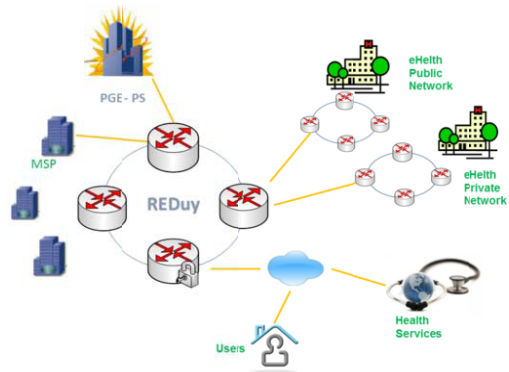


Figure 1 – The eHealth network

Table 1 – Context Summary

Context Summary
The Nationwide Integrated Healthcare System establishes the right to quality healthcare, individual's right to access his or her own health records and that this medical information should be available when needed.
The EHR is a key component for the Nationwide Integrated Healthcare System success.
All health institutions that are part of the health system must contribute to the HCEN.
Each citizen is affiliated to at least one comprehensive healthcare provider.
Each individual may receive medical care from another institution, different from the healthcare provider to which is affiliated.
Each healthcare provider develops its own EHR.
Many healthcare institutions are using specific applications, e.g., RIS-PACS, LIS.
Some pilot applications on imaging (remote diagnosis) and oncology are being implemented by salud.uy program.
There are some interoperable applications: vaccines, births and deaths certificates, childhood and adolescence cards, old people cards, pregnancies, which are centrally managed by MSP.
Master Tables, Catalogs, Dictionaries are managed by MSP but the e-services provider is the salud.uy program through the eHealth platform.
Terminology services will be managed by salud.uy program. Salud.uy program also will provide them as centralized services.
Currently, each healthcare provider adopts its own criteria for patients and EHR identifications.
There is a base platform, PGE, which was designed considering health standards, and it is currently operating.
Interoperability: a successful model on complexity abstraction has been developed to facilitate the adoption of standards, as well as to ease interoperability itself.
The key components of the health data transmission network are available (health network).

eHealth Platform: Architecture of a Solution at a National Level

Background and Preliminary Studies

During preliminary studies before designing the PS, various models were analyzed, in particular, models from Canada [10], Catalonia, Italy, France, Austria and Australia [11]. Many of their components were considered when designing our model. Of course, we also considered HL7, IHE, ISO, IHTSDO, SNOMED CT, DICOM, LOINC standards and recommendations.

Architecture of a Solution at a National Level

Every ICT platform requires the prior solution of some basic infrastructure functionalities, such as: i) security (users authentication and access control); ii) routing and transformation of messages (ESB); iii) implementation of standards on interoperability and usability; iv) offering continuation mechanisms and data access; v) offering ready access to information (portal); vi) connectivity at different levels; vii) management and administration tools; viii) governance; ix) ensuring privacy and data protection.

All those basic functionalities, including citizen single Sign On, are provided by the PGE. Hence, eHealth Platform (PS) will be considered as a specialized extension of the PGE.

PGE Technical Features

The e-Government Platform has three layers: one is physical and two are logic. The physical layer is formed by the Cloud, which supports the whole platform infrastructure. This physical layer implements the following: Infrastructure as a Service (IaaS) and Platform as a Service (PaaS).

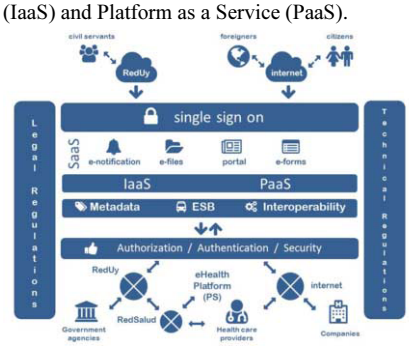


Figure 2 – PGE diagram

This allows selecting the best way to manage the solution infrastructure, optimizing physical resources.

The logic layers, the Crossed-Application layer (Saas) and the Interoperability Platform, refer to two visions of this business: one is oriented to final users (Crossed-Applications layer) and the other is oriented to integration services for backend systems (Interoperability Platform).

Crossed Applications

The layer of Crossed Applications is divided in two layers; security and applications. The security layer, in this case, offers security services not only for internal applications, e.g., State Portal and Electronic File, but for the implementation of a Single Sign On solution which allows adding applications hosted within other State agencies or the PS. This implementation is based on the fact that the identities of application users are federated and that authentication is centralized in the PGE. This is achieved by the use of SAML 2.0 tokens, which allow validating individual's identity through the addition of tokens to HTTP requests.

Interoperability Platform (PDI in Spanish)

The PDI integrates the systems within the State at the back end level [9]. It has two layers: one of security and the other of interoperability (semantic and technical). Technical interoperability is solved by the use of an Enterprise Service Bus (ESB) and the semantic is solved by a number of metadata definitions, which allow simplifying data exchange and offering added value services on it. Those exchanges are based on Web Services SOAP 1.1. All web services used by this platform comply with WS-Security Profile 1.0 and WS-Basic Profile 1.1 standards. SSL v3.0 (HTTPS) with mutual authentication used for physical security. Service authentication and authorization is carried out by SAML1.1 tokens, which are exchanged with the platform using WS-Trust and WS-Security standards. E-signature is used to ensure tokens authenticity through certificates X.509 v3. Messages received by PDI are dynamically routed to various recipients using WS-Addressing standard. Internally, XACML 2.0 is used for validation and communication of security components. XLST 2.0 is used for ESB transformations. Open standards allow universal use of the platform, becoming independent from proprietary protocols and overcoming difficulties at the integration stage.

eHealth Platform (PS in Spanish)

Featured by the existence of multiple healthcare providers, the SNIS creates the need of building a federated model of Electronic Health Records [7]. SNIS requires that everyone is responsible for managing and providing custody for health records of their affiliates, apart from the need of having a complete view of the EHR whenever and wherever it is required.

HCEN context requires the fulfillment of interoperability rules and data integrity rules. First rules are supported by the three main levels of interoperability: organizational interoperability, semantic interoperability and technical interoperability. Data integrity is based on rules and policies on auditing, traceability, identity, reliability and auto-verification.

Organizational interoperability is based on two elements: design based on the affinity domain concept [8] and SNIS regulations.

Designs are based on the affinity domain concept, according to the definition included in IHE technical specification. IHE refers to a number of healthcare providers that share some policies within a cooperation environment, using a shared infrastructure.

Considering that this model was designed for a nationwide system, governance of this domain is required, in order to establish and ensure shared rules or policies with which healthcare providers agree to comply.. Some of the policies to be defined are the following: organizational rules (legal, roles, and structural, basically established by SNIS), operational rules (e.g., SLA, backup and recovery policies, modification and deletion of medical and administrative data), domain registration rules, authentication and access policies, privacy policies, consent, auditing.

Semantic interoperability is based on the definition and management of document formats accepted by repositories, as well as on terms and vocabulary. Moreover, it is also based on the implementation of terminology services, according to a controlled framework with limited and distributed access

Technical interoperability is based on the adoption of some standards and instruments provided by the PDI that are

expanded to include defined standards and profiles established by HL7 and IHE, and implemented on specific instruments such as ESB-PS, EMPI, XDS, and Appliance-PS which are described below.

The integration engine (ESB-PS) is formed by the ESB from the Interoperability Platform, to process specific messages of health protocols. It plays the role of transforming, validating, routing, integrating and performing broadcast actions on messages received from health applications. Within this engine, various source and target ports are set up to carry out data exchange processes. Messages are issued and sent from and to the applications that will be integrated, according to different communication protocols (TCP/IP, HTTP, SOAP, WS*, REST, MLLP) and different messaging formats (HL7 V2.x, HL7 V3, DICOM, X12, XML, EbXML, EDI, delimited text). After receiving the message from a port, it is identified and processed syntactically. Messages are directed within the engine and may be validated and/or transformed. After message is transformed, it may be sent to another target port, according to certain format and protocol. As any ESB, it provides functionalities for managing and monitoring service and traffic.

Appliance-PS allows easy interoperability between healthcare providers and the PS. It is based on lessons learned on e-Government interoperability and it seeks an important level of abstraction to simplify messages exchange and integration. It is an artifact formed by hardware and software that implements layers of services and simple user interfaces. It has two components: one is located in the PS (software) and the other is physical (software and hardware) and is located in provider's sites on a high availability basis. The main functionalities of this artifact are:

Locally, for each provider: i) PGE connector, which solves connectivity, authentication and access to PGE and PS; ii) Messaging, registry XDS (IHE) and HL7v2 and HL7 v3; iii) To make an easier use of EMPI through abstraction and encapsulation of services; iv) Messaging services for XDS Registry for healthcare providers which do not have XDS, as well as for specialized repositories, e.g., PACS and LIS, no XDS. v) Messages withholding in case of interruption of services; vi) Management of Publish & Subscribe services to update EHR of their patients and affiliates; viii) Functionalities for local use of digital signature.

Regarding PS: i) Transforming to CDA Salud.uy those documents which contain minimum required data, but do not meet format standards; ii) specific adapters according to non-standard system's requirements; iii) other nonstandard transformation service for legacy systems.

Other Specialized Components of the PS

The PS hosts the National Index of Healthcare Users (INUS in Spanish), usually known as master table of patients, enterprise version (EMPI), which contains updated information of different individuals ID to supply the affinity domain. This allows each provider to maintain its current ID and to share medical information with the HCEN through a local XDS registry, managed by the provider, which is copied to the National Registry of Medical Information. This registry is also located within the PS, and it receives the information supplied by local registries kept by each provider.

To keep the EHR updated, and in order to minimize data collection during patient's visit, healthcare providers have a Publish & Subscribe service. They may subscribe to this service in order to receive reports on medical acts occurred within the system related to their affiliates. Furthermore, healthcare providers may subscribe to receive medical information related to users registered on special vertical programs, e.g., chronic diseases.

The PS also contains necessary software components to deliver MSP main services related to: i) individuals health; ii) healthcare services (master tables); iii) dictionaries and terminology; iv) reporting systems to inform MSP.

Authentication and authorization

The authentication and authorization model of PS is based on the PGE model, which is based on the federation of identities and, hence, on reliability. This model is centered on PGE and uses SAML 1.1 tokens signed with X509 v3 certificates, to be the unique access point to different services. To do this, the PGE uses a system based on RBAC, where every provider and service consumer has a branch in the LDAP directory of the PGE. In this directory, organization existing roles are defined, which will be used afterwards to authorize the consumption of certain service. This authorization may be achieved with finer granularity: to the extent of a web service operation. Specific roles used by health applications will be managed within the respective applications.

Integrating Multiple Repositories - Specialized and Distributed

Reality demonstrates that different healthcare providers, as well as the PS, have specialized services which require specialized repositories; e.g., PACS (local for each provider and centralized, e.g., imaging experimental plan), LIS, HCEO and other highly specialized services. According to reality, those repositories not always comply with XDS standards; this is the case of current PACS and LIS. But regarding HCEN, it is necessary to integrate information and access. The goal is to obtain an integral view of the patient's EHR, regardless used technologies or data location.

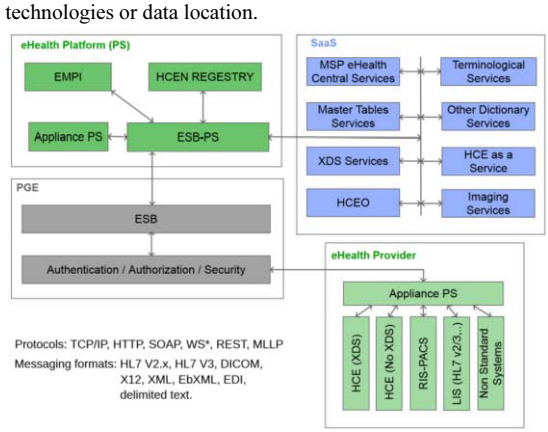


Figure 3. e-HEALTH Platform - components view

Prerequisites: 1) All medical acts, regardless the act type, must be included in the Registry of the HR; 2) All documents, images, reports, among other documents, must be available to duly authorized healthcare staff, whenever is required; 3) The final format of representation must comply with document standards defined by Salud.uy; specifically, regarding images it will be established suitable compression levels for each intended use; 4) Documents must be electronically signed by the healthcare provider, to ensure the identity, authenticity and inalterability of the document; 5) Local representation adopted by each provider regarding their own repositories must ensure the existence of the minimum data required for each kind of document. Moreover, each provider must ensure the fulfillment of the above-mentioned prerequisites.

Mechanisms for the integration of repositories: This diagram (see Figure 3), which naturally shows a diversity of technologies, requires the existence of an integration layer, capable of routing and message transformation. Data transformation to standard formats of presentation is also required. This layer may have specialized services, e.g., Dicom images and their services of publication and access. Figure 3 shows a simplified view of its implementation. As it is illustrated, Appliance-PS is a key part of this architecture.

Current Stage of the PS Development

PGE platform is working with very high and increasing rates of message traffic and implemented services per day. All the functionalities mentioned herein, which are required for the PS operation, are currently operating in a daily basis.

PS platform is still under construction. However, there are two experimental applications which are already operating: HCEO and remote imaging diagnosis. Both led to the experimental use of certain tools, e.g. OpenEmpi, MIRTH. Moreover, it was developed as a specific XDS, which gave us relevant experience and learning. Currently, pilot activities on terminology services are being carried out, in cooperation with the Italian Hospital from Buenos Aires (HIBA in Spanish). Uruguay has entered into an agreement with the International Health Terminology Standards Development Organization (IHTSDO), and plays an active role as an IHTSDO member.

Evolution of the PS platform

This platform will be further developed by the acquisition, construction and integration of products. In addition, the development and creation of tables, dictionaries, minimum data sets, MSP regulations, among other required elements, will contribute to its continuous evolution. Each healthcare provider shall continue with the development of their own EHR, according to standards and technical specifications laid down. Moreover, an appliance-PS will be supplied, which should be integrated in their respective platforms. Cloud services (SaaS) will be further developed in order to make EHR registration, search and exchange of information easier for small health care institutions.

Table 2 – Platform Summary

Platform Summary
The PS is a specialized extension of the PGE, which uses it as foundation for common functionalities.
PGE´s architecture is defined in three layers: a Physical layer, implementing IaaS and PaaS, and two logical layers, IaaS or Crossed-Aplication and the Interoperability Platform (PDI).
All three interoperability layers are covered (Organizational, Semantic and Technical) by the eHealth Platform.
Healthcare providers may be in different stages of technical development. The use of the Appliance-PS accompanied with the PGE Connector fulfills the technical gap, simplifying the interaction with the PS, as well as giving a powerful tool for service governance.
Each Healthcare provider will maintain their user registry, keeping their own ID and information, but feeding the National Index of Healthcare Users.
The existence of a local XDS provided with the Appliance-PS, allows to share medical information with the HCEN
Even the PS is still in working progress, HCEO and remote imaging diagnosis application are on pilot over it.

Conclusion

It has been established that a technological platform with a scalable, reliable, and secure architecture fulfills the needs of the Nationwide Health System. The platform also considers the public policies focused on citizens, contributing to improve their quality of life, and duly assuring information security and personal data protection.

Lessons learned on interoperability at a national level have been applied to the health platform. The integration between the e-government platform and the e-health platform, as well as the operation of the referred networks—salud.uy, red.uy and private networks—allowed the reuse of knowledge, infrastructure and services.

Complex engineering mechanisms related to authentication and access control, exchange of messages assuring security, privacy and information completeness, as well as supporting various protocols and data models, has been solved and will be used by all health care providers.

The challenge that lies ahead is the incorporation of different participants to the eHealth Network, as well as achieving citizens' ownership and use of this platform.

References

[1] AGESIC. Digital Agenda Uruguay 2011- 2015. http://www.agesic.gub.uy/innovaportal/v/1443/1/agesic/mapa_de_ruta:_agenda_digital_uruguay_2011-2015.html?menuderecho=11

[2] Republic of Uruguay, Legislative Branch. Act 18.211, Nationwide Integrated Healthcare System (SNIS). http://www.parlamento.gub.uy/leyes/AccesoTextoLey.asp?Ley=18211&Anchor=

[3] Republic of Uruguay, Legislative Branch. Act 18.335, HealthCare Patients and Users, Rights and Duties. http://www.parlamento.gub.uy/leyes/AccesoTextoLey.asp?Ley=18335&Anchor=

[4] AGESIC. Technical standards and best practices. http://www.agesic.gub.uy/innovaportal/v/114/1/agesic/Normas-Tecnicas.html

[5] Object Identifier (OID) National Managing Body (UnaOid). http://unaoid.gub.uy/

[6] AGESIC. PGE-Connector. http://www.agesic.gub.uy/innovaportal/v/2817/11/agesic/conector_pge.html

[7] ISO Standard. Federated health record13606-1:2008 https://www.iso.org/obp/ui/#iso:std:40784:en

[8] IHE. IT Infrastructure Technical Framework - ITI TF-1:10. http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol1.pdf

[9] AGESIC. PDI User guide. http://www.agesic.gub.uy/innovaportal/v/3543/9/agesic/guia-de-uso-.html

[10] INFOWAY. https://infowayconnects.infoway-inforoute.ca/index.php/resources/technical-documents/architecture

[11] NEHTA. Personally Controlled Electronic Health Record (PCEHR). https://www.nehta.gov.au/implementation-resources

Address for correspondence

Jorge Abin: jorge.abin@agesic.gub.uy; jabin@fing.edu.uy

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-17

Electronic Dental Records System Adoption

Renata Abramovicz-Finkelsztain[a], Claudia G N Barsottini[a], Heimar Fatima Marin[a]

[a] *Department of Health Informatics, Universidade Federal de São Paulo, Brazil*

Abstract

The use of Electronic Dental Records (EDRs) and management software has become more frequent, following the increase in prevelance of new technologies and computers in dental offices. The purpose of this study is to identify and evaluate the use of EDRs by the dental community in the São Paulo city area. A quantitative case study was performed using a survey on the phone. A total of 54 offices were contacted and only one declinedparticipation in this study. Only one office did not have a computer. EDRs were used in 28 offices and only four were paperless. The lack of studies in this area suggests the need for more usability and implementation studies on EDRs so that we can improve EDR adoption by the dental community.

Keywords:
Electronic Dental Records; adoption; case study.

Introduction

During recent years, there has been an increase in the adoption of current generation technologies in dental offices. Some reasons for adoption include: increasing "green" practices, increasing efficiency, usage of technology as a marketing element, and exchanging patient information with insurance companies.

Green practices include the use of digital photography and radiographs as well as the use of EDR to decrease the consumption of paper and the production of chemical waste.

To improve efficiency and productivity, some offices use rotary instruments, apex locators, computerized anesthesia, intraoral cameras and scanners and CAD-CAM systems.

Additionally, the use of any of the above mentioned technologies themselves are perceived as marketing elements for the patient. The patient notices the up to date context of the office and services provided.

The presence of a computer in an office also allows usage of other applications such as patient scheduling, communication with patients, colleagues and providers, clinical and financial management and also the use of EDR systems.

Lastly, some offices have acquired computers to be able to exchange information with insurance companies, as these companies continually demand this exchanged information to be made digital.

Literature Review

As highlighted by the literature, the use of technologies in dental education began in 2000. In clinical practice computers, in particular, have already been used as an administrative management tool since the 1980's [1]. In 1996, Paul R. Rhodes discussed the main differences between electronic and paper records, as well as their advantages and disadvantages [2]. The visual resources of the EDR were the most positive features in the author's opinion because once the patient was allowed to visualize graphically his or her clinical needs, the satisfaction and acceptance of the treatment were enhanced [2]. Visual resources were also seen as important features by Delrose and Steinberg for use in patient education [1].

As more technologies for dentistry became available on the market, research to evaluate the relationship between dentists and these technologies began to evolve. In Canada there was research to evaluate dentists' perception regarding the use of new technologies and also to determine the presence and use of computers in Canadian dental offices. Researchers verified that 60% of the dentists believed that technology could improve their clinical practice and 90% of the dentists already had computers in their offices [3, 4]. Similar results were found in the USA and England, with 85% of American dentists having computers in their offices while 77% of British dentists having or declaring the intention of acquiring a computer in the near future [5, 6].

Having established the presence of computers in the dental offices, and the use or intention of using an EDR system, authors began to investigate the annotation fields both in paper and electronic records. They noticed a great difference between the records and reasoned that the digital records have a limited coverage of the patient clinical information. Furthermore, they found that fields that are usually together in paper records, are often separated in the EDRs, possibly making it difficult to be filled by the dentists [7]. In order to investigate possible usability problems on EDR systems, Thyvalikakath et al carried out a heuristic evaluation of four systems [8]. The authors described 229 heuristic violations and suggested potential usability problems in all four systems [8]. Despite the deficiencies and difficulties already described, EDR systems are commonly used in dental offices in North America and Europe, as well as in some North American dental schools. Studies have also been performed in dental schools. These studies cover the use of EDRs for a wide variety of themes: pharmacology education [9], the use of controlled terminology to annotate diagnosis [10] and treatment planning by undergraduate dentistry students [11].

A literature review conducted in 2014 showed that despite the great number of articles regarding EDR published over the last decade, only 22% of those were about EDR design and architecture, while 78% were editorials, reviews or articles describing the use of EDR data. In addition, taking a closer look at the articles regarding EDR, only 20% of the studies represented research related to EDR use (adoption, usability, implementation, tools). The other 80% was research that used clinical data from the EDRs to make new scientific discoveries.

Objectives

This study aims to identify and evaluate the adoption of EDR systems in the São Paulo city area.

Methods

Due to the fact that the study comprises contemporary phenomena, a case study design with a quantitative approach was the chosen method for this research [12, 13].

The Ethical Committee of Universidade Federal de São Paulo approved this project.

First, a translation, adaptation and validation of the questionnair,used by Schleyer et al [5] in a similar study in 2006, was performed. This occured after getting permission from the mentioned author. This instrument was chosen for the study because it is well established and frequently used in the literature.

A dentist with proficiency in English did the translation and adaptation into Portuguese. In order to validate the translation, another dentist with proficiency in English back-translated the questionnaire into English. Both versions were compared and if there were a disagreements between the translations these were reconciled by both dentists [14].

A convenience sample of dentists working in the Sao Paulo city area was selected.

Due to the length of the questionnaire (31 questions) and the resistance of the dentists to spend 20 minutes over the phone answering it, the questionnaire was divided into two sections. The first section contained questions regarding the presence of computer and internet access, the size of the staff, quantity and location of computers; and therefore, could be answered by any staff member (receptionist, dental assistant, dental hygienist or the dentist himself). The second section, with more specific details over technologies and EDR adoption was shorter and more objective in order to be answered by the dentist himself.

Each office was contacted at least twice: on first contact the purpose of the study was explained and if consent to participate in the study was given, the first section of the questionnaire was answered. A second telephone contact was made at a more convenient time for the dentist to answer the second section of the questionnaire.

Results

Up to the present moment, seventy-one offices were contacted. For 28 offices there were no successful contacts and for the other 53, contacts were successful.

Of the 54 contacted offices, only one did not have a computer. Of the 53 computerized offices, three did not have internet access, and 47 had cable internet access. In two of the offices that did not have internet access, the computer was used rarely and at the third one daily. For the rest of the offices, the computer was used daily at 44 offices and weekly at the remaining four (Figures 1 and 2).

In terms of the location of computers: four offices had computers only the reception, ten offices had computers only in the office, 11 offices had computers in the reception and in the office, 13 offices had computers in the operating room and in the reception and 15 offices had computers only in the operating room (Figure 3).

The use of EDR systems was observed in 28 offices. The most commonly used system was Easy Dental™ (Easy Software SA, Brasil), being used in 19 offices. The software DentalMaster™ (Micro Imagem, Brasil) was used in four offices and Excel™ (Microsoft, USA) was used in two offices (Figure 4). The majority of offices kept only paper records (29 offices) and only six offices were paperless. The remaining 17 offices maintained records both in paper and in digital format. (Figure 5)

In the offices contacted, there was a great variation regarding the insertion of information in the records as shown in Figure 6.

Figure 7 shows the frequency of computer and internet use to search clinical information.

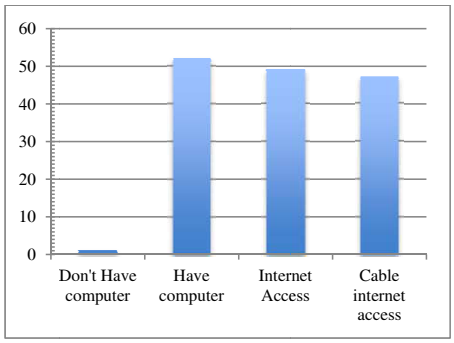


Figure 1- Presence of computers and internet access

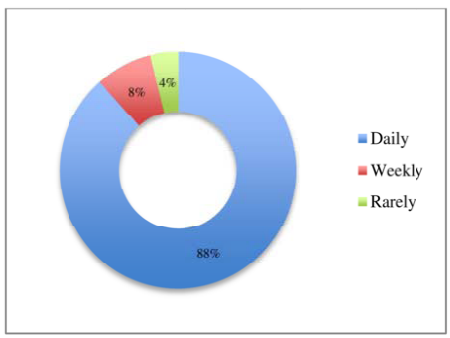


Figure 2- Computer use frequency

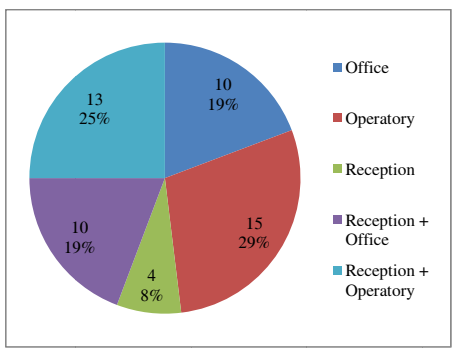


Figure 3- Location of computers

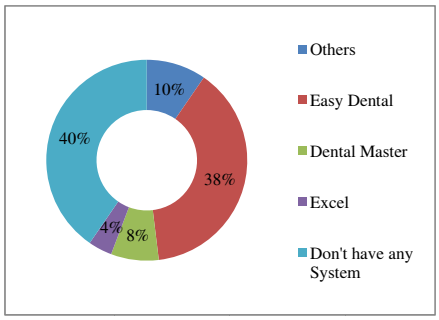


Figure 4- Adopted EDR Systems

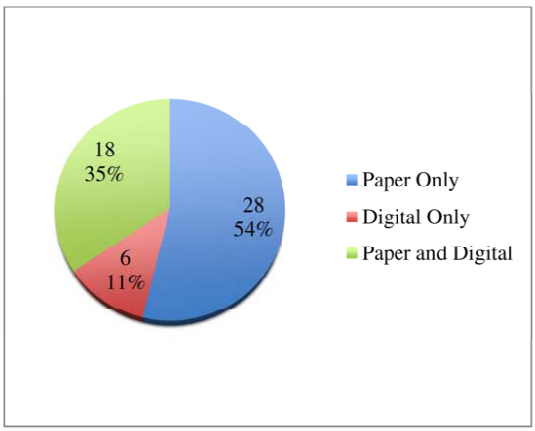


Figure 5- Dental records format

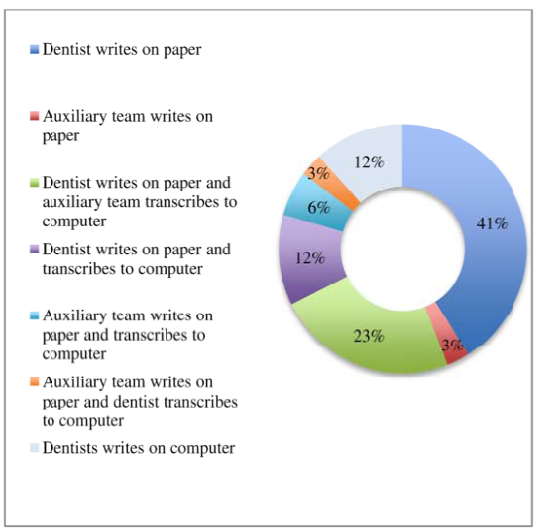


Figure 6- Insertion of information on dental records

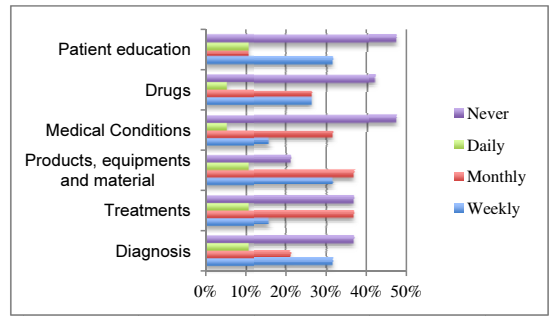


Figure 7- Clinical information search using internet and computers

Discussion

Although computers are available in the majority of the contacted offices (98%), EDR systems have not been used by most of the contacted dentists. The most common scenario is the duplication of records (paper and digital), resulting in an increase in the time spent writing notes on records and a decrease in productivity [5]. Some of the reasons for this duplication of records might be: difficulties in using the systems, mistrust for the EDR systems and a lack of a standardized nomenclature in dentistry. In Brazil, there are at least 14 available EDR systems, with some dating back to 1994. Even though they have been available for 20 years, none of the systems complies with the security requirements of the current Brazilian legislation. In addition, neither the federal, nor the regional Dentistry Councils issue digital certificates for the professionals.

The lack of a standardized nomenclature has been addressed in a number of studies, and specifically in Brazil, the nomenclature used dates from 1995 [15]. More recently, in 2007, the Brazilian National Health Agency established a nomenclature pattern to be used in the information exchange by all insurance companies, including dental ones. Since then, most EDR systems use this pattern to describe dental treatments, but still misses a standard for dental diagnosis. It is important to stress that only the dentists that are registered with the insurance companies use this nomenclature, and therefore, the other professionals are not familiar with it.

Another reason for the incomplete adoption of the EDR systems might be the fact that some of the dentists declared being forced into adopting the systems by the insurance companies, since processing issue treatment authorizations have to be done online. Therefore, they use only the functionalities that are necessary for the authorization process, i.e., registering the patient, his/her insurance information and the suggested treatments.

Although present in the majority of the contacted offices, Internet and computers are not usually used to obtain clinical information. Almost 50% of the contacted dentists never use them to obtain information on medical conditions, and almost 40% never use them to obtain information on diagnosis. Nevertheless, these results are similar to those found in other international studies [3, 5]. In our study, most offices (73%) had two or more dentists, therefore having more than one specialty in the clinic enabling them to consult with one another for this type of information.

People are not afraid of technology anymore. New generations grow up in a world where technology does not intimidate and the way we produce and store information needs to be

efficient [16]. In addition, patients subconsciously register that better dentists have tech savvy offices; while those that do not adopt technology tend to be considered less competent [17].

Even though technology is usually charged with being responsible for impersonal relationships, professionals using them are more productive. This leaves more available time for dentists to bond with the patient in order to strengthen their relationship, resulting in greater acceptance of the suggested treatments and a more satisfied patient.

Even though the sample of this study is small, and not representative of the entire country's dental community, it should influence local councils to issue more comprehensive surveys in order to describe the regional scenario with more details. This way, policies and incentives might better influence the EDR systems adoption by the dental community.

Conclusions

This case study demonstrates a high level of modernization in dental offices as well as adoption of EDR systems, but little clinical use. Contacts with the dentists are still being made until the end of January 2015.

References

[1] Delrose DC, Steinberg RW. The clinical significance of the digital patient record. Journal of the American Dental Association (1939). 2000;131 Suppl:57s-60s.

[2] Rhodes PR. The computer-based oral health record. Journal of dental education. 1996;60(1):14-8.

[3] Flores-Mir C, Palmer NG, Northcott HC, Huston C, Major PW. Computer and Internet usage by Canadian dentists. Journal (Canadian Dental Association). 2006;72(2):145.

[4] Flores-Mir C, Palmer NG, Northcott HC, Khurshed F, Major PW. Perceptions and attitudes of Canadian dentists toward digital and electronic technologies. Journal (Canadian Dental Association). 2006;72(3):243.

[5] Schleyer TK, Thyvalikakath TP, Spallek H, Torres-Urquidy MH, Hernandez P, Yuhaniak J. Clinical computing in general dentistry. Journal of the American Medical Informatics Association : JAMIA. 2006;13(3):344-52.

[6] John JH, Thomas D, Richards D. Questionnaire survey on the use of computerisation in dental practices across the Thames Valley Region. British dental journal. 2003;195(10):585-90; discussion 79.

[7] Schleyer T, Spallek H, Hernandez P. A qualitative investigation of the content of dental paper-based and computer-based patient record formats. Journal of the American Medical Informatics Association : JAMIA. 2007;14(4):515-26.

[8] Thyvalikakath TP, Monaco V, Thambuganipalle HB, Schleyer T. A usability evaluation of four commercial dental computer-based patient record systems. Journal of the American Dental Association (1939). 2008;139(12):1632-42.

[9] Gregson KS, Romito LM. Using the patient's medication history as a learning tool in clinical pharmacology instruction for dental students. Journal of dental education. 2012;76(11):1482-90.

[10] Tokede O, White J, Stark PC, Vaderhobli R, Walji MF, Ramoni R, et al. Assessing use of a standardized dental diagnostic terminology in an electronic health record. Journal of dental education. 2013;77(1):24-36.

[11] Tokede O, Walji M, Ramoni R, White JM, Schoonheim-Klein M, Kimmes NS, et al. Treatment planning in dentistry using an electronic health record: implications for undergraduate education. European journal of dental education : official journal of the Association for Dental Education in Europe. 2013;17(1):e34-43.

[12] Yin RK. Estudo de Caso: Planejamento e método. 2ºed. Porto Alegre: Bookman; 2001.

[13] Hancock DR, Algozzine B. Doing case study research: A practical guide for beginning researchers: Teachers College Press; 2006.

[14] Alves FS, Pinto Rde M, Mendonca TM, Silva CH. [Portuguese-language translation and cross-cultural adaptation of the Fatigue domain of Patient-Reported-Outcomes Measurement Information System (PROMIS)]. Cadernos de saude publica. 2014;30(5):1103-10.

[15] Organization WH. Application of the international classification of diseases to dentistry and stomatology : ICD-DA - See more at: http://apps.who.int/iris/handle/10665/40919 - sthash.GcusBLqr.dpuf: Geneva: World Health Organization; 1995. 3rd edition:[Available from: http://apps.who.int/iris/handle/10665/40919.

[16] Tapscott D. Growing up digital: McGraw-Hill New York; 1998.

[17] Sam FE, Bonnick AM. Office computer systems for the dental office. Dental clinics of North America. 2011;55(3):549-57, ix.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-21

Characterizing the Structure of a Patient's Care Team through Electronic Encounter Data Analysis

Shan He[a], Greg Gurr[b], Susan Rea[a], Sidney N. Thornton[a, c]

[a] *Intermountain Healthcare, Salt Lake City, Utah, United States*
[b] *Agilex, Chantilly, VA, United States*
[c] *University of Utah, Salt Lake City, Utah, United States*

Abstract

As the field of medicine grows more complicated and doctors become more specialized in a particular field, the number of healthcare providers involved in healing an individual patient increases. This is particularly true of patients with multiple chronic conditions. Establishing effective communications among the care providers becomes critical to facilitate care coordination and more efficient resource use, which will ultimately result in health outcome improvement. The first step for care coordination is to understand who have been involved in a patient's care and their relationships with the patient. The widespread adoption of Electronic Health Records provides us an opportunity to explore solutions to well-coordinated care. This paper presents the concept of a patient's care team and demonstrates the feasibility of identifying relevant healthcare providers for an individual patient by leveraging electronic patient encounter data. Combined with network analysis techniques, we further visualize the care team structure with quantified strength of patient-provider relationships. Our work is foundational to the larger goal of patient-centered, coordinated care in the digital age of healthcare.

Keywords:

Care coordination; Health information exchange; Electronic Health Record; Patient care team; Network analysis.

Introduction

Care coordination has been identified by the Institute of Medicine (IOM) as one of the priority areas for health care quality improvement [1]. According to IOM, the aim of care coordination is to "establish and support a ***continuous healing relationship***, enabled by an integrated clinical environment." The Agency for Healthcare Research and Quality (AHRQ) defined care coordination as "the deliberate organization of patient care activities between two or more participants (including the patient) involved in a patient's care to facilitate the appropriate delivery of health care services,…, and often managed by the ***exchange of information among participants*** responsible for different aspects of care" [2]. Improved care coordination would have an especially important impact on improving health care processes and outcomes for patients, particularly for those with one or more chronic conditions [3,4]. It also has great potential in promoting efficient use of limited resources and cost savings by avoiding duplicate tests and unnecessary hospitalizations. A recent report published by the Center for Disease Control and Prevention (CDC) estimated that approximately half (49.8%) of US adults have at least one chronic condition and about 1 in 4 adults has multiple chronic conditions (MCC) [5].

Patients with chronic conditions are usually seen by many providers at several locations such as family medicine clinics, specialty care clinics, hospitals, birth centers, urgent care facilities or emergency rooms. The average Medicare beneficiary sees 6.4 different physicians in a year, 4.6 of those being in the outpatient setting [1]. Without well-designed care coordination among various care providers, disconnects in services can occur and result in adverse drug events, missed follow-ups, unnecessary hospitalizations, and duplicative tests.

In order to achieve well-coordinated care, effective communications among the care team members and patients need to be supported. However, the first step for enabling effective communication is to identify who the team members are and what type of treatment relationship each member has with the patient. Certain providers such as primary care physicians (PCP) may maintain an ongoing relationship with the patient that lasts for years or even decades. Some providers may have a continuous relationship with the patient for a period of time. For example, patients usually go to the same obstetrician/gynecologist or maternal fetal medicine (MFM) doctor for a series of prenatal visits. However, the patient-provider relationship usually ends when the perinatal care ends. Other providers may only have a visit-specific relationship with the patient, such as emergency or inpatient physicians. Understanding the characteristics of patient-provider relationships is critical to implement standard, evidence-based care process models and deliver relevant, timely, and accurate patient information to the right care providers, patients and their families. Stage 2 of the EHR Incentive Program overseen by the Centers for Medicare and Medicaid Services (CMS) in the United States requires the summary care record for each transition of care to include a list of care team members beyond the transitioning provider and the receiving provider [6].

Related Research

Electronic Health Record (EHR) systems have been widely adopted in the United States, and a substantial amount of administrative and clinical information have been captured. Such information can be valuable to facilitate care process and improve outcomes if used properly. A previous study evaluated the feasibility of using system audit log data of EHR chart accesses to identity a patient's care team [7]. This study focused on the inpatient setting for a single hospital stay. There is a lack of research in identifying a patient's care team in the ambulatory care settings, which involve a larger number of provider roles and multiple encounters at different facilities. We searched PubMed [8] using MeSH terms "Electronic Health Records" and "Patient Care Team" and did not find relevant articles in the results.

Furthermore, not every provider involved in a patient's care has a continuous relationship with the patient. The closeness

of the relationship between providers and patients varies depending on the patient's conditions and care needs. Network analysis has been widely used in other domains such as biology, geography, web technologies and communication studies. It is a powerful tool to characterize social relationships in terms of network theory, which consists of nodes and ties. Nodes represent individual entities in the network, whereas ties represent relationships between entities [9,10]. However, network analysis has not been widely applied in the healthcare domain.

Project scope

The goal of this paper is to examine the feasibility of leveraging patients' encounter data, which include both clinical and administrative information about each patient visit, to identify patients' relevant care team members. We further quantify and visualize the strength of the relationship between each care team member and an individual patient based on the frequency and timing of office visits using network analysis technologies.

Methods

Healthcare Setting

Intermountain Healthcare is a non-profit system that operates 23 hospitals, a medical group with more than 185 physician clinics, 3,000 affiliated physicians, and an affiliated health insurance company, SelectHealth. It is the largest integrated system of health care facilities in the Intermountain West region of the United States. Intermountain Healthcare serves about 2.6 million patients, with more than 3.7 million ambulatory visits and 400 thousand inpatient stays each year.

Data Sources

The enterprise data warehouse (EDW) at Intermountain Healthcare integrates clinical data, financial data, claims and eligibility data from electronic systems across the entire enterprise [11]. The EDW loads data daily from various source systems such as the clinical data repository, pharmacy, cancer registry, enterprise master patient index (EMPI), accounts receivable, SelectHealth claims, and laboratory and radiology systems. It is designed to support Intermountain's clinical operations and research efforts. The EDW is organized as a collection of various data marts, which consolidate and integrate data from separate systems to support a specific clinical operation area.

We identified data marts in the EDW relevant to the purpose of this project containing patient encounter data dating back to 1997. Although originally designed suitable for specific use (billing), in general the data marts include visit type (e.g. inpatient, outpatient hospital, office visit, emergency department, urgent care, etc.), visit or service date, provider information, diagnosis, facility information, payer, and other financial information. The hospital data mart includes 44 million records and the ambulatory data mart includes 66 million records. Each record has a unique identifier for that specific encounter, which links to the unique patient identifier maintained by the EMPI.

Relational database tables in these data marts were queried using Oracle SQL Developer to extract patient encounter records meeting certain criteria. Resulted datasets were used for further statistical and network analysis. Specifically, queries were developed to address the following questions:

1. How many encounters (both inpatient and ambulatory) are recorded in EDW for each patient? How many encounters does each patient have for the past 7 years, and 3 years?
2. How many care providers are documented in each patient's encounter records? What are their roles?
3. What is the closeness (or distance) of the relationship between each care provider and the patient? Do patients see some care providers more often than others?

These questions are critical to help us understand the feasibility of using encounter data to extract patients' care team. If the number of encounters for each patient is too large, the care team extraction may need to be deployed offline in a batch mode so that it will not impact the real-time user experience. In order to get the complete and most relevant care team member, a time constraint may need to be applied to the encounter records. For example, inpatient stays happened 20 years ago may not have any impact on patient's current care team.

Statistical and Network Analysis

Datasets extracted from Oracle queries were analyzed using R (version 3.1). Descriptive statistics were used to characterize the patient encounter and care team patterns, both inpatient and outpatient. We performed network analysis using Pajek [12] to visualize the structure of a patient's care team. In this paper, we primarily focused on ambulatory visits that may result in a continuous relationship between the provider and the patient. For example, PCP office visits are more likely to indicate a continuous relationship between the provider and the patient compared to inpatient and urgent care visits. This assumption is based on the ultimate goal of this project, which is to facilitate care coordination by notifying a relevant healthcare provider about the patient's condition. Specifically, we extracted encounters in the ambulatory data mart that have a place of service code with any of the following values: office (11), home (12), assisted living facility (13), place of employment-worksite (18), birthing center (25), nursing facility (32), hospice (34), community mental health center (53), or end-stage renal disease treatment facility (65). The place of service codes are standardized by CMS for use on professional claims to specify the entity where service(s) were rendered [13].

As a proof of concept, we randomly selected a few sample patients to illustrate the structure of a care team. The "strength" of the relationship between a patient and each care team member is quantified by taking into account both the frequency and timing of interactions (encounters) between them. For example, if a patient had 20 encounters with his/her family doctor and only 2 encounters with a cardiologist, then we consider the patient has a closer relationship with his/her family doctor. If a patient had 20 encounters with family medicine doctor A, which all occurred five years ago and 7 encounters with family medicine doctor B in the past two years, then we consider the patient has a closer relationship with doctor B. As shown by Equation (1), the strength of the patient-provider relationship (SPPR) represents the closeness between a patient and a provider. It is measured by the total number of visits divided by the number of months between the current date and the last visit date. If the most recent visit is within a month, the closeness is simply measured by doubling

$$SPPR = \begin{cases} total\ number\ of\ visits \times \frac{30}{current\ date - last\ visit\ date} & (if\ current\ date - last\ visit\ date > 30) \\ total\ number\ of\ visits \times 2 & (if\ current\ date - last\ visit\ date \leq 30) \end{cases} \quad (1)$$

the number of visits (as is the case when current date-last visit date =15).

Results

Patient Encounter Patterns

Table 1 summarizes the distribution of the number of both inpatient and outpatient encounters for all Intermountain patients. It is a skewed, long-tail distribution, which can be indicated by the small median values. Although the maximum number of encounters is large, 96% of the patients have less than 100 encounters in the database.

Table1– Summary statistics for number of patients' encounters (derived from both inpatient and outpatient visits)

	Patient Population	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
All data	4.2 M	1	2	7	22.5	25	2278
Last 7 years	2.8 M	1	2	7	17.1	20	1135
Last 3 years	2.1 M	1	2	5	10.9	13	578

Figure 1 shows the number of visits by year for three randomly selected patients. As shown in the diagram, the number of visits across patients can vary greatly and there is no consistent pattern for each individual patient. The frequency of visits mostly depends upon individual patient's health status and disease progress.

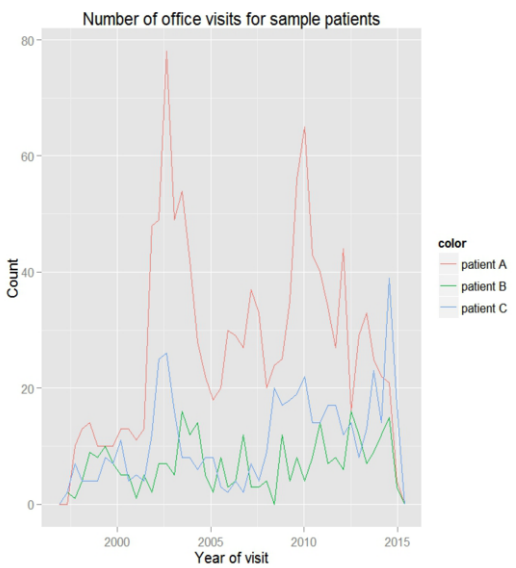


Figure 1– Number of office visits for sample patients

Care Team Patterns

Table 2 summarizes the distribution of the number of unique providers in a patients' care team, identified by querying encounter records for all ambulatory visits. Similar to the encounter distribution, the size of a patient's care team follows a long-tail distribution (as shown in Figure 2). Although the maximum number of care team members for a patient is large, 90% of the patients have less than 10 care providers recorded in their office encounter records. Figure 3 illustrates the distribution of the number of providers for patients with more than 25 care team members.

Table2– Summary statistics for the size of patients' care team (derived from office visits)

	Patient Population	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
All data	2.5 M	0	1	3	4.2	5	93
Last 7 years	1.7 M	1	1	2	3.1	4	56
Last 3 years	1.2 M	1	1	2	2.4	3	34

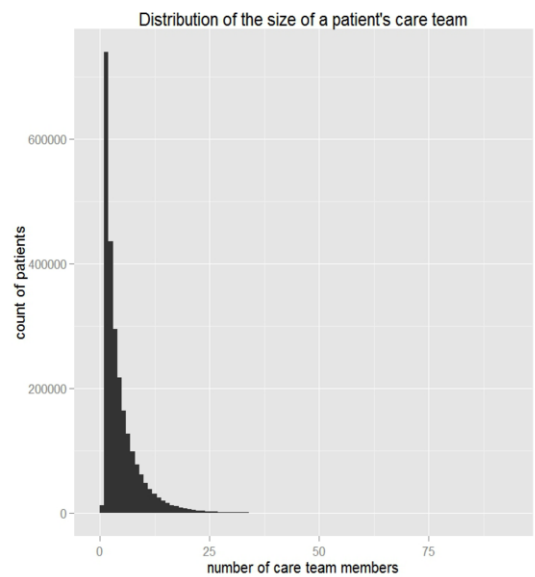


Figure 2– Frequency distribution of the size of patient's care team

Care Team Structure

Figure 4 illustrates the structure for a sample patient's care team. The patient is represented by the red dot in the middle, and each care team member is represented by a green dot labeled by a number index with his/her specialty. The distance between red dot and each green dot represent the strength of the patient-provider relationship, which is calculated as described in the method section. As a proof of concept, only a subset of the sample patient's care team is included in the diagram.

Discussion

Care teams were be built for most patients in real-time. For a small percentage of patients with large number of encounters, care team were be built offline to alleviate the impact on system

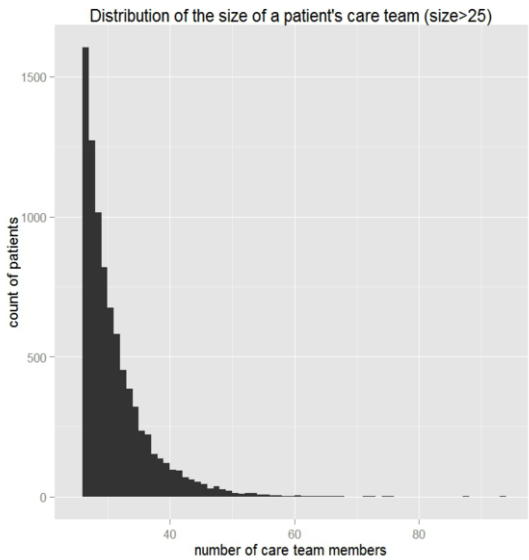


Figure 3– A zoom-in view for the distribution of the size of patient's care team for patients with more than 25 unique providers

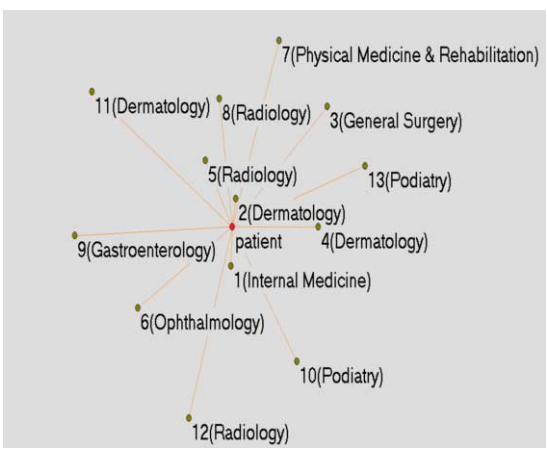


Figure 4– Structure of a sample patient's care team

performance. The structure of a patient's care team evolves dynamically and should be updated each time a patient has an encounter.

To generalize the results to other institutions, careful considerations should be followed to identify the most appropriate data sources containing healthcare providers' information.

The study is limited by the lack of formal evaluation of our proposed measurement for the patient-provider relationship. By manually examining sample patient records, the measurement reflects the closeness between a patient and healthcare providers. However, a structured methodology needs to be employed to validate and improve the accuracy of our proposed measurement. A possible approach would be to display the care team structure to patients and clinicians in patient portals or provider portals and capture users' feedback.

For future work, we plan to extend the care team analysis by including documents received from health information exchange transactions. The sender information indicates a treatment relationship with the patient and our extended analysis will generate a more comprehensive care team by including providers outside of Intermountain. In addition, we plan to analyze and visualize the provider-to-provider relationships beyond the patient-provider relationships (e.g. a cardiologist and the referral doctor). This is important to identify the disconnections in the patient care process. We plan to explore the association between the number of diseases of a patient and the size of his/her care team. Finally, to validate and improve the accuracy and completeness of the derived care team, we plan to display the care team structure to patients and clinicians in patient portals or provider portals and capture users' feedback.

Conclusion

Curating the list of the healthcare providers from multiple care settings who are involved in a patient's care to keep them informed with relevant patient health information is critical to achieve effective care coordination and improve health outcomes. We demonstrated the feasibility of identifying patients' care team members using the electronic patient encounter records.

References

[1] Adams K, and Corrigan JM. Priority Areas for National Action: Transforming Health Care Quality. 2003.

[2] McDonald KM, Sundaram V, Bravata DM, Lewis R, Lin N, Kraft SA, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 7: Care Coordination) [Internet]. Agency for Healthcare Research and Quality (US); [cited 2014 Dec 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK44015/

[3] Anderson G, and Knickman JR. Changing the chronic care system to meet people's needs. Health Aff (Millwood) 2001:20(6): 146-60.

[4] Wagner EH. The role of patient care teams in chronic disease management. BMJ 2000: 320(Feb): 569–72.

[5] Ward BW, Schiller JS, and Goodman RA. Multiple Chronic Conditions Among US Adults: A 2012 Update, 2014; pp. 4–7.

[6] Centers for Medicare & Medicaid Services. 42 CFR Parts 412, 413, and 495 Medicare and Medicaid Programs; Electronic Health Record Incentive Program—Stage 2. Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology 2012: 77(171): 1–196.

[7] Vawdrey DK, Wilcox LG, Collins S, Feiner S, Mamykina O, Stein DM, et al. Awareness of the Care Team in Electronic Health Records. Appl Clin Inform 2011: 2: 395–405.

[8] PubMed [Internet]. [cited 2014 Oct 2]. Available from: http://www.ncbi.nlm.nih.gov/pubmed

[9] Scott J. Social Network Analysis (third edition). SAGE Publications Ltd, 2012.

[10] Abraham A, and Hassanien A-E. Computational Social Network Analysis: Trends, Tools and Research Advances. Springer, 2009.

[11] Evans RS, Lloyd JF, and Pierce LA. Clinical use of an enterprise data warehouse. AMIA Annu Symp Proc 2012; pp. 189–98.

[12] Pajek Wiki [Internet]. [cited 2014 Dec 22] Available from: http://pajek.imfm.si/doku.php.

[13] Place of Service Codes for Professional Claims (updated November 1, 2012) [Internet]. 2012. [cited 2014 Dec 22] Available from: http://www.cms.gov/Medicare/Coding/place-of-service-codes/Place_of_Service_Code_Set.html

Address for correspondence

Shan He shan.he@imail.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-26

An Information Paradigm Shift is Required to Realize EHR Benefits

Evelyn Hovenga[a], Heather Grain[a]

[a] *EHealth Education Pty Ltd, Melbourne, Australia – www.ehe.edu.au*

Abstract

The use of EHRs and the benefits from them are significant for healthcare and health informatics. Data form the basis for any EHR and its potential to realize these benefits. This paper considers the housing of information and knowledge management in an EHR system. Are we experiencing a revolution in healthcare? Findings from an investigation of alternative approaches, followed by an evaluation of the importance of the adoption of a standard information model relative to benefit realization, is presented. We conclude that an EHR in any environment is not limited to sharing or information exchange. A paradigm shift in thinking, based on the requirement for standardized concept representation, is required. This is an essential prerequisite for a new vision of healthcare supported by digital technologies.

Keywords:

Electronic Health Record, Knowledge Management, Terminology, Medical Informatics, Education, Technical Standards

Introduction

Healthcare in its current form is not sustainable [1]. Information technology is increasingly used as a means to improve information availability and flow to support healthcare delivery and outcome improvement. Despite this, these systems often deliver minimal benefits rather than an integrated raft of positive outcomes for the investment. Consideration of maximizing benefits is rarely undertaken.

Healthcare is dependent upon information systems from disparate vendors who have their own information structure, which they value as intellectual property. It is to the advantage of vendors to retain their existing structure as this makes their systems unique and encourages take up of associated products from the same vendor. The impact of this is information representation that is inconsistent between products and systems. There is a need for the introduction of a disruptive technology.

Healthcare is a team activity requiring multiple uses of the same data by different members of the team at different times for a variety of purposes. They need to make use of numerous alternative technologies even though many information system implementations focus on singular activities, or worse singular data representation minimizing multiple data use opportunities.

Traditional knowledge acquisition methods, such as clinical trials, are limited in scope, costly to run, and take time to deliver viable results. Information stored in health records could be used to reduce time and costs associated with knowledge discovery but this is not on the agenda for most healthcare system implementations. Traditional reporting and statistical analysis for public health, finance, and other purposes are equally expensive and dedicated to singular purposes, rather than finding their place in the data continuum.

A limiting factor in decision making and project management in health informatics is a lack of understanding on the part of healthcare administration. That is, many healthcare administrators neither know essential components, nor do they possess the skills needed to support their decision making and projects. There is a resistance to start fresh with systems due to existing investment, yet when major new investments are made the long-term vision is often lacking.

Transition strategies to move towards realizing a long-term vision are not in place. Such transition requires the use of existing systems in a manner that enables a progressive move towards a defined vision. Lack of understanding of such a journey is a consistent problem, as demonstrated by a focus on projects of implementation, rather than of progressive development of data, systems and people.

Methods

This research represents findings based upon a range of information gathering methodologies. These include literature review, participation in, and interviews with, key experienced eHealth standards developers and implementers, and an analysis of:

- The outcomes of eHealth initiatives and implementations
- The capabilities of current systems and approaches
- The outcomes declared
- Skill development initiatives
- The characteristics of data, information and knowledge and the relationship between the information to be shared, stored and retreived and how that relates to the technical infrastructure, and
- Inhibitors/enablers such as funding arrangements, organisational structures.

This paper focuses on an evaluation of our findings. It's important to differentiate between EHRs and EHR Systems. EHRs are essentially a data repository that needs to be an integral part of an entire healthcare system. Early information systems in healthcare were all about reporting and finance. Today these requirements, though still present vie with the need to represent data for healthcare requiring different data representations. This variety of data representation in information collection systems and the requirements for standardized but different representations for different healthcare reporting and other purposes present a unique data dilemma.

These differences stem from two primary causes:

1. The requirement for data use in patient care to accurately and reliably represent facts,

2. Reporting systems requiring classified data that aggregate concepts.

Classified data indicate where necessary that a case fits in the 'other category' or special rules apply for how to collect the data (e.g., when coding clinical data into the International Classification of Diseases (ICD), any version codes). Each approach to data is relevant to purpose but every purpose requires different data presentations whilst data also needs to be managed across the continuum of use. Though the principle of data re-use is often discussed, the need for consistent approaches to achieve the conversion of data from disparate systems and information models is a challenge. This challenge is made more difficult as the original data meaning may vary based on context. These differences lack clarity in the absence of a comparable information model in a system using meaning based data representation.

The variations in original systems and the requirement needs for standardized reporting, mean that conversion approaches such as data maps are being increasingly used. Such maps are only consistent when built and applied to a single known purpose [2] and must be kept up to date. Lack of appreciation of the implications of using maps results in data which appear comparable but which may not actually be so.

Importance of a Standard Information Model

Health software vendors have traditionally controlled the structure and representation of information in healthcare systems and consider these data structures proprietary. This has resulted in data representation and systems that do not share common meaning and require effort to share data with others and to manage data over time. This approach discourages healthcare system purchasers to change their systems to reduce disruption and cost, but this approach exponentially increases costs due to increasing maintenance costs with potentially negative impacts on safety, and the ability to cost effectively, accurately and comprehensively represent EHR data across the healthcare continuum.

Investigation of implementations of electronic health record systems to date has shown that it is common for these proprietary systems to be chosen and implemented with a vision aimed at solving simple problems, such as transitioning from paper based to electronic record systems or to support improved information access with little consideration of the greater list of functional requirements [3-5]. Nguyen et al. [3] identified serious 'concerns regarding the accuracy and completeness of records'. These authors also identified a need for further work into information system quality for EHR implementation evaluative research. The result of this extensive literature has demonstrated that current EHR limitations have resulted (or will result) in the need to change systems. Furthermore, systems must undergo costly reinvention of the data, and associated data exchange protocols, for current systems to be extended to address new functional needs as these arise.

The research gap identified also demonstrates a lack of understanding of the new paradigm of health data pertaining to the world healthcare initiative's focus on information exchange. Though there is no doubt such exchange is necessary, a more critical requirement is to be able to accurately extract subject data via computer processing without loss of meaning. There is a strong relationship between data accuracy and the technical schema used for data exchange. This is poorly understood and rarely seriously considered in system implementation, yet functional requirements determine the degree and type of interoperability and hence technical schema/system architecture required.

Electronic Health Record systems are being widely adopted with the intention of delivering some or all of the following benefits:

- Longitudinal patient records – records able to be retained, queried and retrieved over time to support patient care and knowledge acquisition.
- Retrieval and presentation of the right information to the right care provider or to the patient/caregiver to support clinical care and improve health outcomes
- Use of clinical decision support to improve the application of clinical knowledge and health outcomes
- Facilitate information exchange
- Enable patient / caregiver access to information
- Support reporting and data re-use
- Enable knowledge acquisition from systems
- Support research and epidemiology/public health

The implementation of electronic health record systems tends not to differentiate between the record and the systems used for data collection and exchange to support such records. Existing systems and paper-based records often convert data without considering the actual clinical knowledge relationships between the data and systems prior to the need for conversion.

In some cases, much has been achieved despite this lack of vision and understanding of system capacity and requirements through the individual efforts of clinical informaticians with the knowledge to creatively implement system approaches to data and knowledge [6]. Where understanding and skills to implement change are combined, significantly more is achieved. Of the implementations around the world, which have delivered significant benefit, few have done so without leveraging such expertise and leadership.

Significance of System Interoperability

Many standards are designed for information exchange for a purpose. HL7 internationally is moving towards a single terminology representation shared across all of its products, and many countries are standardizing national data dictionaries. The move to the use of well managed terminologies such as SNOMED CT, LOINC or the use of a machine readable terminology source such as the Unified Medical Language System (UMLS), support this need. These innovations are not required, however, to be applied to health records *per se*. We have found through experience that a general reusable terminology cannot serve all aspirations for clinical information systems that need to make use of data contained in EHRs. In addition, it is necessary to consider data structures and the meaning of data (context) relative to the technical system architectures in use. Only then are distributed systems enabled to exchange information in a meaningful and accurate manner. This requires system architectures that make use of a common standard reference model.

The development and use of sustainable clinical information (concept) models, used in conjunction with these terminologies, that use standard data types and defer to a standard information model, is less well understood and rarely implemented, Brazil [7] and Norway are two notable exceptions. National adoption is the best possible solution to maximizing the value of EHR adoption for all parties, including software developers/vendors.

Semantic requirements between distributed systems as used by the health industry were investigated with a goal of determining whether, and perhaps why, healthcare differs from other industries with similar issues.

Health data is certainly more complex than other industries (e.g. accounting), but the complexity, or even the dynamic nature of health information, is not that which sets it apart from other industries. Rather, it is that this complex and frequent evolution of knowledge must be continually accommodated by EHRs in a timely manner.

Accountancy has used a consistent information model for many hundreds of years. The terminology used is largely numeric and therefore consistent and comparable. The general ledger approach is consistent around the world and understood by all systems, with local modifications within known rules and knowledge of concepts, such as income, profit and loss, or cost of goods sold. Healthcare does not have such a standard structure for information and this, along with the complexity of health data, is a significant hindrance to progress.

Adoption of a standard information model provides context. It supports simpler and cheaper information exchanges between multiple systems using the same information model [8]. It also supports the development and maintenance of rule-based clinical decision support, which can be applied universally in such systems, automation of reporting and many other healthcare functions requiring extensive accurate data and knowledge use. This includes accurate data aggregation from multiple individual EHRs (big data) and linkages with other types of data for research, public health, and epidemiological use.

Why is This Not Well Understood?

The literature regarding research aimed to achieve semantic interoperability for distributed systems has by and large been undertaken by software architecture developers and other technical experts, who as a rule do not appear to have fully appreciated or understood health data characteristics, its variability, evolution, uses, or the need for accuracy.

A review of the courses offered for health professionals revealed that it does not have a workforce with an understanding of data and systems, today's methods of data collection, or health record data storage structures. Projects associated with health data collection rarely include an education or skill development strategy for the data, information, knowledge continuum nor its link with available information and communication technologies. Often, existing expertise is not sought, and implementation plans are developed from scratch. In addition, a review of ICT courses offered by Australian Universities revealed that semantic interoperability is rarely mentioned, nor is data science a topic that is routinely included in such courses.

The lack of a skilled workforce and poor use of expertise is another contributing factor to the cost of new initiatives and change. The health workforce is large and highly skilled already. The inclusion of additional knowledge in already stressed courses is problematic unless this can be integrated in current curricula. A significant limitation is that those who teach in university programs are often not cognizant of the impact of, or need for, knowledge of health informatics in their professions [9].

Health Informatics as a profession has worked to define its body of knowledge and to encourage quality education [10]. However, the existing workforce needs practical and often just-in-time learning opportunities. Only a small number are likely to return to university-based education to develop the skills they need. Australian universities offering Health Informatics courses struggle to attract and sustain a significant number of students. Alternative approaches to skill development are needed.

Another reality is that educational organizations have few health informatics skills. The research-based university model delivers highly skilled individuals with highly specialized knowledge, but few with skills in teaching the broader emerging societal and industry requirements of health informatics. Maintenance of skills across this broad area of needed knowledge is difficult and undervalued. Universities are hampered by the lack of demand, employers do not understand the skills they need in their organizations, and the workforce does not see rewards for having knowledge of health informatics. This cycle of demand must be broken if investment in education of practical value is to be delivered.

Discussion

If the electronic health record is to be sustainable, a technology agnostic solution is needed. An effective solution would not be system-dependent, but rather it would be data-dependent. Only then is the delivery of health record systems that are vendor, time and technology independent, possible. If this is achieved, patients, providers, vendors, organizations and governments can move to new technologies, and continue to use and gain value from the data in existing systems. It will be possible to develop any number of niche application systems that all link to EHRs with increasing sophistication.

A nationally shared standard information model, which is clinically valid, maintained and represents an ontology based concept representation system, is a key requirement to achieving this sustainability. CEN/ISO 13606, a European norm also approved as an international ISO standard [11], identifies a common high level model and archetype (content) models for such information. It defines rigorous and stable information architecture, designed to achieve semantic interoperability in electronic health record communication [12]. Implementations of this approach include the OpenEHR clinical knowledge resources, which represent a maximal model of the concepts stored in healthcare records [13]. Content models using standard terminologies such as SNOMED CT or ICNP to represent clinical concepts can reference this information model.

The relationship between the information model, concept representation and the technical infrastructure or systems architecture needs to be better understood. Collectively these components deliver digital systems that suit all healthcare environments. Achieving this vision requires the implementation and use of internationally accepted formal and informal (e.g., well governed open source standards). Adoption requires a deliberate initiative aimed at taking advantage of new technologies. This transitioning to a fully digital world is no different than transitioning from the stone or industrial or information ages; It's about empowering a paradigm shift to the use of a disruptive technology. For such a shift in thinking to be achieved there is a need for relevant professional development of the health workforce. What is the knowledge and skill gap?

Unless we understand and leverage the paradigm shift we will be implementing solutions which are not sustainable, which require change and are expensive to constantly fiddle with. The health and ICT workforce need to appreciate that:

- Change regarding data use is with us (from a technology perspective, who knows what is next).
- There is need to adopt strategic thinking, to better plan, take advantage of opportunities and minimise risk.
- Decision-making methods need to change; implementation decision making needs to be in line

with a vision for the electronic record, beyond that of simply implementing a system, or replacing paper based records. That vision seeks a more long-term sustainable solution. Community and workforce expectations can be met (sustaining the workforce) via integrated just-in-time learning.

Knowledge gaps include:

- Understanding, using and governing data, including the technical infrastructure needed to maximise potential use and the continuum of data use).
- Ability to identify key data for which data governance strategies need to be adopted.
- Understanding the pathway from business case and functional specifications development of EHRs and related system, to system acquisition, implementation and use - this is a journey; purchasing software is not a destination.
- The knowledge and skills needed to best complement existing expertise and meet potential new role requirements.
- Knowledge and skills required to design and implement systems that meet patient/customer, organisational, disciplinary, optimum work and information flow and national data requirements to suit multiple purposes.

A complete understanding of what is required of data and systems to achieve the desired outcomes across our data and systems is needed. Such understanding enables a vision of the future and a truly progressive transition to that future.

Conclusion

It was found that projects and systems around the world tend to focus on delivering local immediate needs at the expense of considering a bigger vision. Future proof and sustainable EHRs with associated systems for which the local implementation is simply a beginning representing a significant component of a much bigger system. It is about small data supporting healthcare at the point of care, but also about big data supporting research and new knowledge acquisition and discovery.

Therefore we concede six unresolved priorities: 1) Adoption of a standard open information model for all of healthcare, 2) use of consistent concept representation where appropriate, 3) an information and knowledge governance strategy and implementation approach, 4) the conversion of existing knowledge into a computational format consistent with the information model, 5) a concept representation system, and 6) appropriate workforce skill development. This includes the automation of reporting or statistical analysis. At the resolution of such priorities, technology will evolve along with the field's knowledge of best practices, and data collection processes will be usable, calculable, and sustainable.

Every nation needs a transition strategy that makes the best possible use of existing EHR systems, concurrently with a future vision and a strategic pathway towards realizing that vision. This requires a paradigm shift in thinking by all decision makers.

References

[1] Coiera E, Hovenga EJS. Building a Sustainable Health System in: IMIA Yearbook of Medical Informatics 2007, Schattauer, Germany.

[2] ISO TR 12300 Mapping of Terminological Resources

[3] Nguyen L, Belluci E, Nguyen LT, Electronic health records implementation: An evaluation of information system impact and contingency factors. International Journal of Medical Informatics Vol.83:11; 2014, pp779-796 DOI: http://dx.doi.org/10.1016/ j.ijmedinf.2014.06.011

[4] Hyppönen H, Saranto K, Vuokko R, Mäkelä-Bengs P, Doupi P, Lindqvist M, Mäkelä M. Impacts of structuring the electronic health record: a systematic review protocol and results of previous reviews. Int J Med Inform. 2014 Mar;83(3):159-69. doi: 10.1016/j.ijmedinf.2013.11.006.

[5] Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform. 2008 May;77(5):291-304.

[6] David V Pilcher, Toni Hoffman, Chris Thomas, David Ernest, Graeme K Hart, Risk-adjusted continuous outcome monitoring with an EWMA chart: could it have detected excess mortality among intensive care patients at Bundaberg Base Hospital. Crit Care Resusc 2010 Mar;12(1):36-41

[7] Brazilian Ministry of Health http://www.openehr.org/news_events/industry_news.php?id=108

[8] Fernández-Breis JT, Maldonado JA, Marcos M, Legaz-García Mdel C, Moner D, Torres-Sospedra J, Esteban-Gil A, Martínez-Salvador B, Robles M. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. J Am Med Inform Assoc. 2013 Dec;20(e2):e288-96. doi: 10.1136/amiajnl-2013-001923.

[9] Wald HS, George P, Reis SP, Taylor JS. Electronic health record training in undergraduate medical education: bridging theory to practice with curricula for empowering patient- and relationship-centered care in the computerized setting. Acad Med. 2014 Mar;89(3):380-6. doi: 10.1097/ACM.0000000000000131.

[10] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, Hovenga E, Lun KC, Marin H, Martin-Sanchez F, Wright G; IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods Inf Med. 2010 Jan 7;49(2):105-120 *http://www.ncbi.nlm.nih.gov/pubmed/20054502*

[11] ISO13606-1:2008 Health Informaics – Electronic health record communication –Part 1 Reference model, retrieved 11 December 2014, available from http://www.iso.org/iso/catalogue_detail.htm?csnumber=40784

[12] The CEN/ISO EN13606 standard, EN13606 Association retrieved 11 December 2014, available from http://www.en13606.org/the-ceniso-en13606-standard

[13] Clinical Knowledge Manager, openEHR Foundation retrieved 10 December 2014 available at http://www.openehr.org/ckm/ & http://www.openehr.org/openehr_programs

Address for correspondence

Dr Evelyn Hovenga, CEO, eHealth Education Pty Ltd, 503/166 Wellington Pde, Melbourne Vic 3002, Australia

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-30

Template and Model Driven Development of Standardized Electronic Health Records

Stefan Kropf[a], Claire Chalopin[a], Kerstin Denecke[a]

[a] *Innovation Center Computer Assisted Surgery, University of Leipzig, Germany*

Abstract

Digital patient modeling targets the integration of distributed patient data into one overarching model. For this integration process, both a theoretical standard-based model and information structures combined with concrete instructions in form of a lightweight development process of single standardized Electronic Health Records (EHRs) are needed. In this paper, we introduce such a process along side a standard-based architecture. It allows the modeling and implementation of EHRs in a lightweight Electronic Health Record System (EHRS) core. The approach is demonstrated and tested by a prototype implementation. The results show that the suggested approach is useful and facilitates the development of standardized EHRSs.

Keywords:

Electronic Health Records, Reference Standards, Information Storage and Retrieval, Medical Informatics Applications, Medical Records Systems, Computerized, Health Information System

Introduction

Patient data are distributed in heterogeneous information systems [1] Accordingly, physicians need to interact with several systems to get a complete view of the health status of a patient. In addition, researchers need to access and reuse patient data for calculations or studies. In both scenarios, it is crucial to have all relevant patient data available and accessible in a standardized, patient-specific information model (IM) [2]. The concept of digital patient modeling adresses this issue and aims at integrating patient data which is distributed in several information systems [3]. Patient modeling on an information model layer corresponds to the development of standardized EHRSs, both for research and for health care, preferably following the open source approach [4]. At a first glance, the development of standardized EHRs and systems that handle these EHRs (an EHRS) seems to be an old-fashioned topic, where practical instructions and solutions already exist. Indeed, EHR standards are available. However, there is a lack of accessible instructions regarding how to apply these standards in practice, and more specifically, how to use standards in real world environments. Most reviewed health informatics-related standards can be classified [5] (ISO 17119) from the perspective of "what," along with conceptual specificity. Standards which can be classified from the alternative perspective of "how," and with *physical* specificity – like the ECG protocoll definition defined in EN 1064 – are rare. This is also reflected by the large number of publications that mainly concentrate on conceptual or logical specificity [6-8]. For implementing standardized EHRSs, developers need methods and instructions or practical solutions like Free Open Source Software (FOSS) analyses [9], the storage of data in persistence layers [10] or ideally an API, [11] development workflows, [12] or even a complete usable EHRS [13]. Therefore, this work introduces a system architecture that combines openEHR and XForms for the development of standard-based EHRs and EHRSs. These standards have been chosen since XForms is a well-established standard of the World Wide Web Consortium (W3C), and openEHR is recommended by e-health authorities such as NEHTA [14]. We used these technologies to build an information system for monitoring patients diagnosed with pituitary adenoma as a proof-of-concept. This paper focuses on the technical details of the development process.

Materials and Methods

We suggest a model-driven architecture based upon existing standards. Such a design principle decouples modeling processes from development processes. Standards are used to fulfill basic requirements in system design such as accessibility, reusability, and interoberability to achieve long term durability. The approach combines W3C-standards originating in web development with openEHR standards (see figure 1). These two groups of standards are interconnected: Instances of openEHR archetypes are mappable to XML.

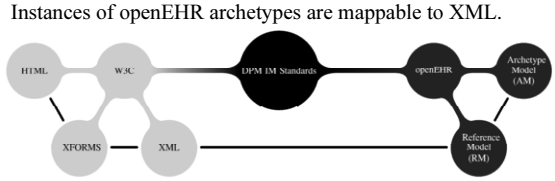


Figure 1: Selection of related W3C and openEHR standards

In the following, the selected standards, methods, and the development process are described. Later, we reflect upon issues that have been considered in the development process.

OpenEHR

Choosing the right standard for an EHRS is a challenge [15, 16]. Normative EHR information models have already been published by the European Committee for Standardisation (CEN)/Technical Committee (TC) 251. These European Norms are based notably on openEHR [17] and Health Level 7 (HL7) v3 Reference Information Model (RIM) [18]. OpenEHR archetypes are transformable to EN 13606 archetypes and vice versa [19, 20]. Both EN 13606 [17] and EN 14822 [18] are suitable for the information model, but openEHR delivers the most comprehensive approach. In short, openEHR delivers a large set of generalized, predefined, standardized building blocks (the archetypes). Reuse of these building blocks is ideal, because they provide a highly

qualitative result of a collaborative development process of domain experts.

However, the collection and ordering of data should be a doctor's first function [21]. openEHR enables domain experts to create information structures without the need of a detailed technical understanding. Such domain modeling by experts is a great opportunity, because physicians can be integrated into the structuring process of medical records, [22] yielding a better understanding of the importance of well-structured, finely granular medical records. Further, it minimizes the need of tasks that invoke expensive downstream processing of natural language.

XForms

XForms is a standard published by the W3C. It follows the Model View Controller (MVC) pattern by seperating the model from its representation [23].

Figure 2 illustrates the XForm technique: The GUI and the input fields are bound directly to the XML via the XForms standard. Important for further reading is the model tag inside the XForms markup. It is located in the header of the file and includes the instance of the model, mapped to XML with possible localized namespace declarations.

The body of the XForm consists of model entities, referenced by a simple "ref" attribute. Figure 2 shows this convenient approach. A simple XPath reference binds input fields to the model definition. This binding concept demonstrates the core idea: The openEHR and XForms standards are glued together to benefit from their respective advantages. The digital patient model of one single EHR is represented by the corresponding model element in an XForm [2-12 in Figure 2].

```
1  <html [...]
2          <xforms:model id="4463197">
3                  <xforms:instance>
4                          <dpmim:myehr [...]
5                                  <dpmim:value>[...]
6                                  <dpmim:magnitude>507.8</dpmim:magnitude>
7                                  [...]
8                                  </dpmim:value>
9                          [...]
10                         </dpmim:myehr>
11                 </xforms:instance>
12         </xforms:model>
13 <xforms:body>[...]
14         <xforms:input ref="/dpmim:myehr/[...]/dpmim:value/dpmim:magnitude">
15                 <xforms:label>Cortisol</xf:label>
16         </xforms:input>
17         [...]
18 </xforms:body>
19 </html>
```

Figure 2: XForms model binding concept

Related OpenEHR Based EHRS

There are several openEHR-based frameworks available. Some of them are independently usable by the underlying persistant layer, such as Chen's openEHR Java reference implementation (http://github.com/wware/openehr-java/). Other approaches are bound to schemes in relational databases. LiU EEE (https://github.com/LiU-IMT/), medrecord (http://www.medrecord.nl/overview/), and EHRflex (http://ehrflex.sourceforge.net/) use XML databases for their persistant layer. Initial tests showed that the Representational State Transfer (REST) architecture of LiU EEE and medrecord is the most promising approach for realizing an EHRS. The REST architecture was already tested in openEHR implementations and showed how an EHRS should work physically [24, 25]. Consequently, we have chosen a REST architecture for our implementation. In addition, our approach combines XForms and openEHR, which results in a standardized storage of patient data in the system. To the best of our knowledge, such an approach has neither been introduced nor implemented previously in the context of patient modeling.

Development Process

The development process is divided into (i) model development and (ii) model-driven application development. This separation of the modeling process from the development process is highly beneficial, as it "empowers the health professional to define and alter the accurate knowledge and information they need in the granularity they need" [22]. Figure 3 shows the separation of work.

Domain Experts develop domain models

Within our concept, domain experts develop the domain model using the openEHR tools archetype editor and template designer. These tools enable clinicians to model their medical records themselves without deeper knowledge about databases.

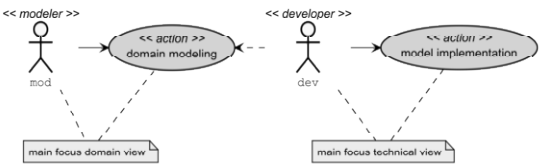


Figure 3: Separation of domain modeling and application development

Figure 4 illustrates the model development portion of the pipeline:

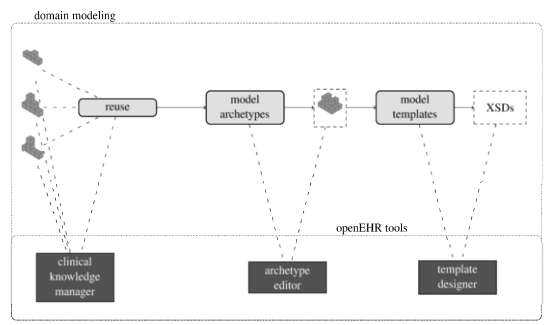


Figure 4: Domain modeling: generation of XSDs underpinned by openEHR tools

The first step in the domain modeling phase is the selection and, thus, reuse of existing building blocks – the archetypes. With the Clinical Knowledge Manager (CKM) [http://www.openehr.org/ckm/], a web repository of openEHR-compliant relevant archetypes for a specific application can be identified. Reuse of archetypes is beneficial for achieving a high degree of semantic interoperability among systems.

After identifying reusable archetypes, they are constrained to the specific use case [26]. New domain-specific archetypes can be modeled with the openEHR-tool archetype editor [27].

This modeling process yields archetypes that are formulated in a standardized language, the Archetype Definition Language (ADL). They are combined in so-called templates (i.e., larger structures), using the openEHR-tool template designer [27]. This concept of building templates out of smaller building blocks increases the probability of reusing single archetypes in different templates. Finally, after this modeling process, models are exported as XML Schema Definitions (XSDs). The W3C XML Schema provides a usable solution for the semantic validation of standard-based EHRs [28]. These

XSDs are then used as input files for the next phase: the model-driven application development.

Model-Driven Application Development and Architecture

Figure 5 shows the process of the application development and the system architecture in general. Out of the XSDs generated in the domain modeling phase, valid XML files are created. We refer to these files by the term "skeleton," because they do not contain real patient data. To put it differently, they are the EHR templates. For each XSD, at least one corresponding XML skeleton is generated. Such an XML/XSD pair enables the validation of the EHRs according to the models at any time, which is highly beneficial, because standard compliance is ensured. The XML/XSD pairs can be exploited as a starting point for development using XML techniques like XSLT, or direct programming of the XML-DOM. However, in this paper, another more accessible solution is proposed: The generated valid XML is inserted into the XForms model and input elements are bound to the XForms model.

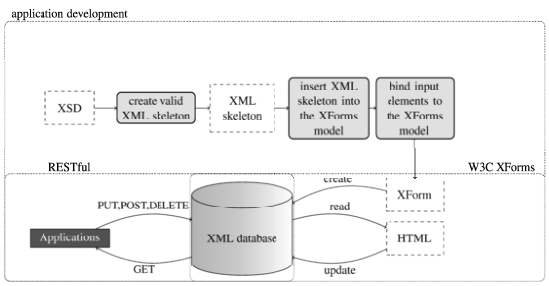


Figure 5: Model driven application development: XForm generation based on XSDs

The source code in Figure 2 shows the approach. To keep this paper as simple as possible, only an extract of one input field is listed. In practice, the instance/model part will be more complex. However, the integration of the model remains simple since it consists of only one copy/paste step. After this insert, the widgets are bound to the model with the "ref" attribute. Finally, the form just has to be designed. After these processing steps, an XForm with an integrated EHR model is stored in the XML database. The form data can then be read and viewed by clients, and updates can be stored and managed by the underlying open souce XML database system eXistdb [29].

In addition to the XML-XForm-GUI binding, the XML database enables the retrieval and storage of complete EHRs via the REST interface, which will be very handy for any kind of future development. In essence, the left side of Figure 5 illustrates that applications can easily access these XML data by a RESTful interface instantly delivered by the XML database eXistdb [30].

Results and Discussion

In this paper, a complete development process for standard-based EHRs was introduced. The domain modeling promises a high quality of the underlying models, and the application development ensures a rapid implementation. To achieve a high level of interoperability, the process starts with the reuse of archetypes, integrates free modeling tools, and ends in an accessible XML-based EHRS. The resulting XML can be interpreted as one possible representation of a personalized digital patient model. The archetypes and templates resulting from the domain modeling task are not only simple structures of information or web masks. Instead, they are valuable sources for other clinicians at other sites, formulated in a standardized language (ADL). These models can be published in worldwide repositories, which allows their reuse in other contexts where they can achieve a similar impact as publications in journals or conferences.

As a result of the combined usage of the EHR standard openEHR and the W3C standard XForms, a high level of accessibility, reusability, interoperability, and therefore long term durability is achieved. Essentially, the combination of the HL7 v3 RIM and XForms would work similar and would benefit also from the following advantages:

Nowadays, most of the information systems are internally based on traditional relational databases [31]. Much effort has to be done to convert this data into different formats (e.g., Extract Transform Load (ETL)). These ETL processes are very time consuming and also error-prone. To avoid these drawbacks, XML-based messages and whole XML-based EHRSs may become a standard solution in medical informatics (at least for small EHRSs).

The benefit of this paradigm shift is obvious when comparing this approach to other developing processes of semantically interoperable EHRs [7]. The main reason is that object-relational class mappings and ETL processes for the generation of messages or single EHRs can be ommitted.

EHRs and messages inside and between hospital information systems (HISs) are defined differently, called the "Message / Record Dichotomy" [6]. With the paradigm shift to XML for EHR storage, it is possible to use EHR extracts stored in data bases directly as messages, and vice versa. The resulting EHRs can easily be transformed to messages for exporting data to another HIS. This is an essential step towards standardized, generalized, and therefore simplified HIS-infrastructures.

Performance

A comparison of querying openEHR based XML databases indicates that XML databases today are not yet the right solution for a persistance layer of big data [32] As a proof-of-concept, we made a load test on the resulting EHRS. 10,000 (64 KB) XML files were transmitted via POST requests, and the time for GET requests was measured. Additionally, the latency time for the XForm-based GUI generation was detected. The results show that the latency ascends constantly with an increasing number of files (see figure 6).

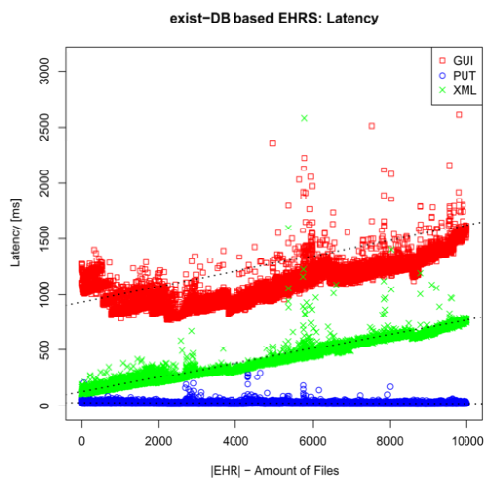


Figure 6: eXistdb based EHRS: Latency

Clearly, the performance of the application slows down with increasing database volume. However, since a total number of 10,000 EHRs will not be reached in our use case, the access times are acceptable. Anyway, because of the ongoing rapid development of storage and processing power, possible performance issues relating larger XML based EHRSs will become more and more negligible.

The Null Flavour Issue

One issue occurring in the XForms solution regards “null flavor” values. To explain this issue, consider the XForm entry in the normal case, e.g. when the value of the element “magnitude” is “507.8”:

<dpmim:value>[…]

<dpmim:magnitude>507.8</dpmim:magnitude>

[…]

</dpmim:value>

The “magnitude” element does not allow empty values, which is of course reasonable. The valid solution to map empty values in XML would be the replacement of the whole <dpmim:value> element with a <dpmim:null_flavour> element, which contains the openEHR code string “271,” which means "no information". Thus, in our example the entire <dpmim:value> element disapears, and is replaced with the <dpmim:null_flavour> element, resulting in:

<dpmim:null_flavour>[…]

<oe:code_string>271</oe:code_string>

[…]

</dpmim:null_flavour>

In combination with the XForm technique, this is an issue, because input fields can not be bound to possibly disappearing elements. One possible workaround is allowing null values in primitive data types: When a value is empty, the relating element will get a null attribute:

<dpmim:magnitude xsi:nil="true"/>

To ensure valid XML, the generated XSD has to be adjusted by allowing null values by the command `nillable="true"` in the correspondent auto-generated XSD element. This workaround allows the use of null values within the XForms model.

Minimal XML

Naturally, XML files come along with relatively much overhead. Additionally, archetypes imply intristic overhead, because they are generalized to enable their usage in different szenarios.

For our prototype system, an XML EHR template containing about 7,000 lines of code for only 24 entry fields was generated out of the XSD files. The overhead was reduced manually by removing unnecessary elements; the final EHR had about a few hundred lines. The important lesson we learned is that contraining the archetype as much as possible is one important prerequisite for the generation of minimal XML-based EHRs.

Conclusions

In this paper, a model-driven and standard-based development process for EHRs has been presented. The practical approach of rapid model driven development of standardized EHRs is recommendable for the development of small EHRSs. Aside from the benefits of standardized EHRs, this boosts the development process. After considering pros and cons of various approaches, a complete paradigm shift to XML for storing EHR data is suggested. XML containing patient data based on openEHR provides a suitable solution for representing digital patient models. The concept was validated by the development of a prototype for a special use case, which is an information system for monitoring patients diagnosed with pituitary adenoma, which will be described in another paper. In future work, additional applications will be implemented that will reuse the concept to use the strength of this approach with respect to application development.

References

[1] Bisbal J, Berry D. An analysis framework for electronic health record systems. Methods Inf Med. 2011;50(2):180--189.

[2] Kropf S, Denecke K. Digital Patient Model - Information Model. University of Leipzig; 2013. p. 123. 12th Leipzig Research Festival for Life Sciences.

[3] Denecke K. Model-based Decision Support: Requirements and Future for its Application in Surgery. Biomedical Engineering/Biomedizinische Technik. 2013;.

[4] Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Methods Inf Med. 2009;48(1):45--54.

[5] Health informatics - Health informatics profiling framework; 2004. ISO/TR 17119:2004(E).

[6] Beale T. Health Information Standards Manifesto. 2001; Available from: http://workflow.healthbase.info/monographs/ [cited 03.11.2014].

[7] Blobel B, Engel K, Pharow P. HL7 Version 3 compared to advanced architecture standards. Methods Inf Med. 2006;45:343--53.

[8] Goossen W, Goossen-Baremans A, van der Zel M. Detailed Clinical Models: A Review. Healthc Inform Res. 2010 Dec;16(4):201--214.

[9] Zuniga F, Enrique A, Win K, Susilo W. Functionalities of free and open electronic health record systems. International journal of technology assessment in health care. 2010;26(4):382--389.

[10] openEHR wiki persistence FAQs;. Available from: http://www.openehr.org/wiki/display/resources/Persistence+FAQs [cited 16.08.2013].

[11] Klasson F, Väyrynen P. Development of an API for creating and editing openEHR archetypes. Linköping; 2009. Masterthesis, urn:nbn:se:liu:diva-17558 (URN).

[12] Lopez DM, Blobel BG. A development framework for semantically interoperable health information systems. Intern. J. of Med. Inform. 2009;78(2):83--103.

[13] Atalag K, Yang HY. From openEHR Domain Models to Advanced User Interfaces: A Case Study in Endoscopy. In: Health Informatics New Zealand Conference; 2010.

[14] NEHTA Review. National E-Health Transition Authority Ltd : Review of Shared Electronic Health Record Standards. 2006. – Research report

[15] Begoyan A. An overview of interoperability standards for electronic health records. In: Proceedings of the 10th International Conference on Integrated Design and Process Technology, IDPT 2007, Antalya, Turkey.2007. p. 3--8

[16] Eichelberg M, Aden T, Riesmeier J, Dogac A, Laleci G. Electronic health record standards--a brief overview. In: Information Processing in the Service of Mankind and Health: ITI 4th International Conference on Information and Communications Technology: Information Technology Institute, Giza, Egypt.; 2006. p. 7--19.

[17] Health informatics - Electronic health record communication [Norm]; 2012. EN ISO 13606 standard familiy.

[18] Health informatics - General purpose information components [Norm]; 2006. DIN EN 14822 standard family.

[19] Martinez-Costa C, Menarguez-Tortosa M, Fernandez-Breis JT. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. J Biomed Inform. 2010 Oct;43(5):736--746.

[20] Costa CM, Menarguez-Tortosa M, Fernandez-Breis JT. Clinical data interoperability based on archetype transformation. J Biomed Inform. 2011 Oct;44(5):869--880.

[21] Weed LL. Medical records, patient care, and medical education. Irish Journal of Medical Science (1926-1967). 1964;39(6):271--282.

[22] Garde S, Knaup P, Hovenga EJ, Heard S. Towards Semantic Interoperability for Electronic Health Records--Domain Knowledge Governance for open EHR Archetypes. Methods Inf Med. 2007;46(3):332--343.

[23] Boyer JM. XForms 1.1 W3C Recommendation [Norm]; 2009. Available from: http://www.w3.org/TR/xforms11/.

[24] Sundvall E, Nyström M, Sandström M, Eneling M, Örman H, Karlsson D. REST Based Services and Storage Interfaces for openEHR Implementations;. Available from: http://www.imt.liu.se/~erisu/2010/EEE-Poster-multipage.pdf [cited 03.11.2014]

[25] Sundvall E, Nyström M, Karlsson D, Eneling M, Chen R, Örman H. Applying representational state transfer (REST) architecture to archetype-based electronic health record systems. BMC medical informatics and decision making. 2013;13(1):57.

[26] Beale T. Archetype Object Model; 2007. Revision: 2.0.1. Available from: http://github.com/openEHR/specifications/raw/Release-1.0.1/publishing/architecture/am/aom.pdf [cited 03.11.2014].

[27] Heard S., Beale T. openEHR – Modeling Tools. Available from: http://www.openehr.org/downloads/modelingtools [cited 03.11.2014].

[28] Rinner C, Janzek-Hawlat S, Sibinovic S, Duftschmid G. Semantic validation of standard-based electronic health record documents with W3C XML schema. Methods of information in medicine. 2010;49(3):271.

[29] Meier W. eXist: An open source native XML database. In: Web, Web-Services, and Database Systems. Springer; 2003. p. 169--183.

[30] Cahlander L, Cross P, Geldiyev Z, Stallman A, Turpin M. eXistdb open source XML database software architecture description. W3C: Introduction to Software Architecture. 2010.

[31] DB-Engines Ranking; Available from: http://db-engines.com/de/ranking [cited 22.04.2014].

[32] Freire SM, Sundvall E, Karlsson D, Lambrix P. Performance of XML Databases for Epidemiological Queries in Archetype-Based EHRs. In: Scand Conf Health Inform. vol. 13; 2012. p. 51--57.

Address for correspondence

Stefan Kropf / Innovation Center Computer Assisted Surgery

Semmelweisstr. 14

D - 04103 Leipzig, Germany

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-35

Fast Model Adaptation for Automated Section Classification in Electronic Medical Records

Jian Ni[a], Brian Delaney[b], Radu Florian[a]

[a] *IBM T. J. Watson Research Center, Yorktown Heights, NY, USA*
[b] *Nuance Communications, Burlington, MA, USA*

Abstract

Medical information extraction is the automatic extraction of structured information from electronic medical records, where such information can be used for improving healthcare processes and medical decision making. In this paper, we study one important medical information extraction task called section classification. The objective of section classification is to automatically identify sections in a medical document and classify them into one of the pre-defined section types. Training section classification models typically requires large amounts of human labeled training data to achieve high accuracy. Annotating institution-specific data, however, can be both expensive and time-consuming; which poses a big hurdle for adapting a section classification model to new medical institutions. In this paper, we apply two advanced machine learning techniques, active learning and distant supervision, to reduce annotation cost and achieve fast model adaptation for automated section classification in electronic medical records. Our experiment results show that active learning reduces the annotation cost and time by more than 50%, and distant supervision can achieve good model accuracy using weakly labeled training data only.

Keywords:
Medical Informatics, Information Extraction, Section Classification, Active Learning, Distant Supervision.

Introduction

Medical records include a rich range of data about individual patients, such as medical and family history, medication and allergies, vital signs, laboratory test results, diagnoses, treatment plans, etc. The huge volumes of data existing in the medical records not only provide great medical research opportunities, but also raise big challenges for processing the data and extracting useful information from it.

An *electronic medical record* (EMR) or *electronic health record* (EHR) system has a number of advantages over traditional paper-based medical record systems. One main advantage is that an EMR system is designed to maintain complete, accurate and up-to-date patient data in electronic form, which can be much more easily and effectively accessed and shared by health professionals as well as patients. This makes automatic information extraction and knowledge discovery from large sets of medical and clinical data possible.

Medical information extraction is an application of the automatic extraction of structured information from unstructured or semi-structured electronic medical records [1]. Such information can be used for enhancing healthcare processes, for example automatically alerting doctors about missing or incomplete data, or automatically generating billing codes from electronic medical records. The extracted information can also be used for improving medical decision making, such as determining the most effective treatment for a specific disease.

In this paper, we study one important medical information extraction task called *section classification*. When dictating, writing or electronically entering medical or clinical documents, clinicians typically organize their narratives into *sections*, which are often labeled with frequently used but non-standardized terms termed *section headers* [2]. Howeverclinicians are not restricted in what they can write as a section header. Many headers are ambiguous (e.g., "SUMMARY:", "HISTORY:", "DISCUSSION:"), and sometimes section headers may not be given, yet are part of the sentence (e.g., "On examination, the patient's blood pressure was 130/80."). The objective of section classification is to automatically identify all the sections in a medical document and classify them into one of the pre-defined section types. Examples of section types include *History of Present Illness*, *Past Medical History*, *Family History*, *Physical Examination*, *Review of Systems*, *Medications*, *Assessment*, and *Plan*.

Section classification is an essential component of a medical information extraction system. It provides important context for other automated information extraction tasks such as word sense disambiguation, mention detection and coding. For example, the acronym PE usually means "Physical Examination" in a *Physical Examination* section while it often means "Pulmonary Embolism" in a *Past Medical History* or *Imaging* section. Similarly, mentions of diseases detected in different sections (e.g., *Past Medical History* section vs. *Family History* section) have different implications depending on who has the disease. For automated coding, procedures detected in some sections (e.g., *Past Surgical History* section and *Plan* section) may not be coded for billing purposes.

The current state of the art section classification models are based on supervised statistical machine learning frameworks [2-4]. Like other statistical information extraction models (e.g., mention detection models, relation extraction models), training section classification models typically requires large amounts of human labeled training data to achieve high accuracy. Given that annotating data in the medical domain needs specially trained human annotators, both time and expense costs can accrue quickly. This poses a big hurdle for

adapting a section classification model trained for one medical institution to a new medical institution.

In this paper, we apply two advanced machine learning techniques, *active learning* and *distant supervision*, to reduce annotation cost and achieve fast model adaptation for automated section classification in electronic medical records. Active learning aims to train a model with high accuracy using minimum annotation effort by selecting the most informative data instances to annotate [5,6]. Distant supervision uses a knowledge base to automatically generate weakly labeled training data [7,8]. We design a heuristic active learning query selection scheme, which reduces the annotation cost and time by more than 50%. In addition, we develop a distant supervision approach that can achieve good model accuracy using only weakly labeled training data.

Materials and Methods

Statistical Section Classification Model

We represent a medical document **x** by a sequence of sentences (x_1, x_2,…, x_n), where x_i is the *i*-th sentence in document **x**. A section classification model assigns each sentence x_i in the document with a section tag y_i, indicating sentence section. Our medical information extraction system has 83 section tags. Examples of some frequently occurred section tags and their counts (based on a corpus of about 8,000 medical documents) are listed in Table 1. Some sections such as *Physical Examination* and *Review of Systems* also contain subsections with finer-grained information, e.g., *Vital Signs*, *Heent*, *Abdomen*, etc. Our system has 26 subsection tags.

Table 1– Examples of Section Tags and Their Counts

Section Tag	Count
Physical_Examination	4,593
History_Of_Present_Illness	4,101
Past_Medical_History	3,666
Medications	3,641
Assessment	3,553
Allergies	3,515
Social_History	3,480
Laboratory_Data	3,331
Review_Of_Systems	3,129
Family_History	2,859
Plan	2,397
Past_Surgical_History	2,037
Description_Of_Procedure	1,552
Hospital_Course	1,536
Chief_Complain	1,479
Imaging	1,254

To capture the discourse structure of the medical documents, we formulate the section classification problem as a *sequence labeling* problem, where the section tags of the sentences in a medical document are predicted jointly. The main advantage of a sequence labeling model is that it can model the dependency of the labels of neighboring classification elements (here, a classification element is a sentence), so it has been widely used in many NLP tasks such as part-of-speech (POS) tagging and named-entity recognition [9,10].

A number of probabilistic graphical models have been applied to tackle sequence labeling problems, including Hidden Markov Models (HMMs) [11], Maximum Entropy Markov Models (MEMMs) [9] and Conditional Random Fields (CRFs) [10]. CRFs have demonstrated good performance for many sequence labeling applications. However, the training of CRFs is slow since the training time grows at least quadratically with the number of labels. On the other hand, while HMMs and MEMMs have linear training time proportional to the the number of labels, they are limited in the types of features they can use.

We applied the *Statistical Information and Relation Extraction* (SIRE) toolkit developed at IBM to train the section classification models [12]. The SIRE sequence labeling model can be viewed as a generalization of MEMMs and it has the following properties: i) fast training time (linear training time in the number of labels as MEMMS and fast convergence rate); ii) allows richer features as CRFs.

The SIRE sequence labeling model uses the following conditional probabilistic model of the section tag sequence **Y** given the input document **X** (both are viewed as random vectors):

$$P(\mathbf{Y}=\mathbf{y}\mid\mathbf{X}=\mathbf{x})=\prod_{i=1}^{n}P(y_i\mid y_{i-1},...,y_{i-o},\mathbf{x}) \quad (1)$$

where each individual conditional distribution has the following exponential parametric form (the form is motivated by the maximum entropy principle [13]):

$$P(y_i\mid y_{i-1},...y_{i-o},\mathbf{x})=\frac{e^{\sum_{k=1}^{K}\lambda_k f_k(y_i,y_{i-1},...,y_{i-o},\mathbf{x})}}{Z_i} \quad (2)$$

f_k's are the feature functions, λ_k's are the weights of the features, and Z_i is the normalization constant which ensures that $\sum_{y_i}P(y_i\mid y_{i-1},...y_{i-o},\mathbf{x})=1$. The model assumes that the tag of a sentence in a medical document depends on the tags of the previous *o* sentences.

Given a set of training documents (where the sentences are labeled with section/subsection tags), we train a statistical section classification model $P_M(\mathbf{Y}\mid\mathbf{X})$ of the form (1) and (2) that maximizes the log likelihood of the training data (plus a regularization term to avoid overfitting of the training data). For any new medical document **x**, we can use the trained model to predict (decode) the most likely section tags of the sentences using the Viterbi algorithm:

$$\mathbf{y}^{*}=\arg\max_{\mathbf{y}}P_M(\mathbf{Y}=\mathbf{y}\mid\mathbf{X}=\mathbf{x}) \quad (3)$$

Active Learning

Active learning is a machine learning technique that aims to train a model with high accuracy using as few human annotated data instances as possible [5,6]. The basic idea of active learning is to iteratively select the most informative instances and use these newly annotated instances to improve the model. This is different from traditional passive machine learning techniques in which a model is trained using whatever annotated instances provided.

Suppose we have a set of human labeled training documents ***L*** from some medical institutions, and we have trained a section classification model using ***L***. To adapt the section classification model to a new medical institution for which we only have a set of unlabeled documents ***U***, a typical active learning algorithm works as follows: A section classification model ***M*** is first trained using the initial set of human labeled

data ***L***. Then the model is used to decode the unlabeled documents ***U***, among which the top *m* most informative documents ***A*** are selected to be annotated by human. A new model is re-trained using both the initially labeled data ***L*** and the newly labeled data ***A***. This process is repeated until some stopping criterion is met (e.g., until certain model accuracy is achieved). The steps are summarized in Algorithm 1.

Algorithm 1: Active Learning for Fast Model Adaptation

Input*:* Labeled data ***L*** (from some old medical institutions), and unlabeled data ***U*** (from a new medical institution)
Repeat*:*

1. Train a model ***M*** using human labeled documents ***L***
2. Use ***M*** to decode the unlabeled documents ***U***
3. Select and remove the top *m* most informative documents ***A*** from ***U***, and ask human annotators to annotate them
4. Add the human labeled data ***A*** to ***L*** and go to step 1

Until some stopping criterion is met

In Algorithm 1, *m* is called the *batch size*. A smaller value of *m* will normally incur less annotation cost (total number of human labeled documents) required to reach a certain model accuracy, however, it will incur more cost (iterations) of re-training the model. In practice, the batch size should be selected to balance the cost of annotation and model training.

A key component of an active learning algorithm is the *query selection scheme* that determines which instances are the most informative. See reference [6] for an excellent survey of different query selection schemes. The *Least-Confidence* (LC) scheme is one of the most widely used query selection scheme for active learning [6,14]. It uses the confidence score of an instance under the current model to measure the "informativeness" of that instance. The idea is that the instances with the lowest confidence scores provide the maximum amount of information, and these instances should be selected to be annotated first, as well as added to the training data to improve the current model.

For our section classification model, the confidence score of an instance (here an instance is a medical document) **x** under the current model P_M is:

$$c(\mathbf{x}) = P_M(\mathbf{Y} = \mathbf{y}^* \mid \mathbf{X} = \mathbf{x}) \quad (4)$$

where **y*** is the most likely section tag sequence for document **x** (see (3)). The LC scheme will select the top *m* documents with the lowest confidence scores.

To better measure the "informativeness" of a medical document, we have designed the following *heuristic query selection scheme*. For each document, we count the number of sentences whose per-sentence confidence score is less than a threshold (e.g., 0.9). Documents with higher counts are more uncertain under the current model and hence provide more information or value for improving the model. Our heuristic query scheme will select the top *m* documents with the highest counts.

Distant Supervision

While a traditional supervised machine learning algorithm relies on human labeled data to train a model, a weakly supervised learning algorithm uses "weakly" labeled data generated by some automatic mechanism to train a model. Weakly supervised learning techniques are often used in applications where unlabeled data are readily available but annotating them by humans is difficult and/or expensive.

Distant supervision is a type of weakly supervised learning technique that uses a *knowledge base* to automatically generate weakly labeled training data [7,8]. For example, in [8] Freebase (a large semantic database containing several thousand relations), distance supervision was used to generate weakly labeled data for training relation extraction models. For our section classification problem, we apply the idea of distant supervision and construct a section header-to-tag map knowledge base to automatically label the medical documents.

In medical documents, a *header* of a sentence is a sequence of words in a special form (e.g., all capitalized, followed by colon) that occurs at the beginning of a sentence, which often indicates the start of a new section. A *header-to-tag map* includes entries of headers and the corresponding section/subsection tag associated with every header. We have developed a rule-based header extractor that can automatically extract the headers of the sentences in medical documents. We can then construct a header-to-tag map either from sentence-level annotated documents or from direct header annotation.

To generate weakly labeled training documents, we apply the header extractor to extract the headers of the sentences in the (unlabeled) medical documents, and the sentences are then labeled sequentially using the header-to-tag map as follows. If a sentence has a header *h* which is in the header-to-tag map, then the sentence will be labeled *tag(h)* where *tag(h)* is the associated section tag of *h* in the map. Otherwise, the sentence will be assigned the same section tag as the previous sentence (because without a header, a sentence is likely to be inside of a section).

Furthermore, if we have the section boundary information, i.e., if we know where a section starts and ends, we can use this information to improve the labeling of the documents. A new section tag will be assigned to a sentence if and only if the sentence has a header, which is in the header-to-tag map, and the sentence is located at the beginning of a section.

Results

Active Learning Experiment Results

We first present the results of applying active learning for fast section classification model adaptation. We use the standard *F_1 score* (or *F-measure*) to measure the performance of a section classification model on an independent test set. The F_1 score is defined as the harmonic mean of *precision* and *recall*:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

Precision is the number of correctly detected sections divided by the number of all detected sections in the test set returned by the model (a section is correctly detected by the model if both the section span and the section tag are correctly identified). For example, if the model detects 100 sections and 90 of them are correct, then the precision of the model is 90/100=0.9. *Recall* is the number of correctly detected sections divided by the number of true sections in the test set. For example, if there are 200 true sections in the test set and the model correctly detects 90 of them, then the recall of the model is 90/200=0.45. Ideally, we want to build a model that has both high precision and recall. While these two metrics are often competing with each other, F_1 score considers both

precision and recall, and is a commonly used metric to measure the overall accuracy of a classification model [15].

We start with a set of 1,615 human labeled training documents from medical institution A. The initial section classification model is trained with these 1,615 documents. When tested on a test set of 409 human labeled documents from medical institution A, the model achieved an F_1 score of 0.921. However, when tested on a test set of 397 human labeled documents from a new medical institution B, the F_1 score was only 0.719. This shows that a good section classification model trained for one medical institution may not achieve good accuracy for a new medical institution without any model adaptation.

To adapt the section classification model to medical instituition B, we experiment with two active learning query selection schemes: the Least-Confidence (LC) scheme and our heuristic scheme. The baseline scheme is random query selection, i.e., passive learning where the training docuemts are randomly selected from the unlabeled document pool. The batch size *m* is chosen to be 50.

The learning curves of different schemes are shown in Figure 1, where *x*-axis is the number of human labeled training documents from medical instituion B (which represents *annotation cost* or *learning cost*), and *y*-axis is the F_1 score (in percentage) of the section classification model on the test set from medical institution B (which represents *learning performance*).

As we can see from Figure 1, compared with random selection, active learning can greatly reduce the annotation cost to achieve a certain accuracy. For example, to reach an F_1 score of 0.89, we need to annotate 400 documents under the heuristic query selection scheme, while we need to annotate 900 documents under the random query selection scheme, so active learning reduces the annotation cost by 55%. The performance of the two active learning query selection schemes is very similar, while the heuristic scheme achieves slightly better accuracy (0.89 F_1 score) than the LC scheme (0.887 F_1 score) with a training size of 400 documents.

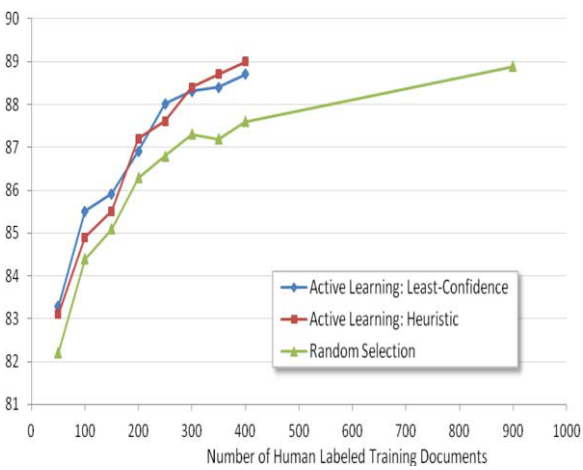


Figure 1– Learning curves of different query selection schemes for section classification model adaptation.

Saving annotation cost means saving of time. A human annotator can annotate approximately 50 documents per day. To reach the same accuracy of 0.89 F_1 score, it would require 18 days to annotate the 900 documents under passive learning, while it would only require 8 days to annotate the 400 documents under active learning, a significant saving of time in adapting the section classification model to the new medical institution.

We have conducted similar experiments to adapt the section classification model to another new medical institution using active learning, and we observed a 57% saving of both annotation cost and time.

Distant Supervision Experiment Results

Next, we present experiment results of our distant supervision approach for fast section classification model adaptation. We start with a set of 3,274 human labeled training documents (denoted as **BASIC**) from three medical institutions A, B, C. We want to adapt the section classification model to a new medical institution D. **NEW** denotes a set of 1,095 medical documents from the new medical institution.

We compare two forms of annotation for the documents in **NEW**: one is human annotation where we use $\mathbf{NEW_{human}}$ to represent the set of training documents labeled by human annotators; the other is distant supervision where we use $\mathbf{NEW_{distant}}$ to represent the set of weakly labeled training documents automatically generated by our distant supervision approach.

As shown in Table 2, a section classification model trained with the data from the old medical institutions (i.e., **BASIC**) only achieves 0.753 F_1 score on a test set of 605 human labeled documents from the new medical institution. This again shows that without model adaptation, a model trained for some medical institutions may not achieve good accuracy for a new medical institution. If we add human labeled training data from the new medical institution (i.e., $\mathbf{NEW_{human}}$), then the model achieves 0.919 F_1 score on the test set. These provide us with two baseline results: the *worst* baseline that does not use any training data from the new medical institution, and the *best* baseline that uses human labeled training data from the new medical institution.

Under distant supervision (as opposed to using human labeled training data from the new medical institution), we use a header-to-tag map to automatically label the **NEW** data. We experiment with two cases. In the first case, we do not have any section boundary information. If we detect a header *h* of a sentence and the header is also in the header-to-tag map, then the sentence will be labeled *tag(h)* where *tag(h)* is the associated section tag of *h* in the map. In this case, the section classification model trained with **BASIC** data and the weakly labeled $\mathbf{NEW_{distant}}$ data achieves 0.902 F_1 score on the test set.

In the second case, the section boundary information is available. Such information can be obtained by direct section boundary annotation, which is much cheaper than complete section annotation since here the annotators only need to label which sentences are located at the beginning of a section and do not need to assign any section tags. With the section boundary information, the performance of the section classification model improves to 0.906 F_1 score on the test set.

Table 2. Experiment Results of Distant Supervision (Test Data are from the New Medical Institution D)

Training Data	F_1 Score
BASIC	0.753
BASIC + $\mathbf{NEW_{human}}$	0.919

BASIC + NEW$_{distant}$	0.902
BASIC + NEW$_{distant}$ (with section boundary)	0.906

To adapt the section classification model to the new medical institution D, under the traditional supervised learning framework (where distant supervision or active learning is not used), it would require approximately 22 days to annotate the 1,095 training documents from medical institution D to get an F_1 score of 0.919.

Under the distant supervision approach, the main annotation cost is the cost to build the section header-to-tag map. The header-to-tag map contains 3,094 headers, among which only 375 headers are new headers from medical institution D. The other 2,719 headers are from medical institutions A, B, C, and the section tags for those 2,719 headers are readily derived from the **BASIC** data set. We only need to ask a human annotator to label the 375 new headers from medical institution D, which can be finished in one day or two – requiring significantly less time than annotating the whole set of 1,095 medical documents from medical institution D. There is a small performance drop of the section classification model trained under the distant supervision approach. Nevertheless, it can still achieve an F_1 score of 0.902, which is good for many applications of interest.

Conclusion

Section classification in electronic medical records provides important contextual information for other automated information extraction tasks. Training section classification models, however, typically requires large amounts of human labeled data to achieve high accuracy and annotating institution-specific data can be both expensive and time-consuming. In this paper, we have applied two advanced machine learning techniques, active learning and distant supervision, to reduce annotation cost and achieve fast section classification model adaptation. In particular, we have designed a heuristic active learning query selection scheme, which reduces the annotation cost and time by more than 50%. We have also developed a distant supervision approach that can achieve good model accuracy using weakly labeled data only.

There are important differences between active learning and distant supervision. Active learning belongs to the family of supervised learning algorithms, which requires human annotators to label the data. By intelligently selecting the data instances to annotate, active learning can achieve the same level of accuracy as traditional supervised learning algorithms with much less annotation cost. In contrast, distant supervision belongs to the family of unsupervised (if no human annotation is used at all) or semi-supervised (if some limited human annotation is used) learning algorithms. Its performance heavily depends on the quality of the knowledge base. Distant supervision normally does not achieve the same level of accuracy as supervised learning algorithms, but it incurs even less annotation cost than active learning. In practice, one should decide which technique to use based on the accuracy requirement and the annotation budget. It is also possible to combine the two techniques for fast model adaption.

References

[1] Meystre SM, Savova GK, Kipper-Schuler KC, and Hurdle JF, Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research, IMIA Yearbook of Medical Informatics, 2008.

[2] Denny JC, Spickard A, Johson KB, Peterson NB, Peterson JF, and Miller RA, Evaluation of a Method to Identify and Categorize Section Headers in Clinical Documents, Journal of the American Medical Informatics Association, 2009.

[3] Li Y, Gorman SL, and Elhadad N, Section Classification in Clinical Notes using Supervised Hidden Markov Model, ACM IHI 2010.

[4] Tepper M, Capurro D, Xia F, Vanderwende L, and Yetisgen-Yildiz M, Statistical Section Segmentation in Free-Text Clinical Records, LREC 2012.

[5] Cohn DA, Ghahramani Z, and Jordan MI, Active Learning with Statistical Models, Journal of Artificail Intelligence Research, 4(1):129-145, 1996.

[6] Settles B. Active Learning Literature Survey. Computer Sciences Technical Report, University of Wisconsin-Madison, 2010.

[7] Craven M and Kumlien J, Constructing Biological Knowledge Bases by Extracting Information from Text Sources, Seventh International Conference on Intelligent Systems for Molecular Biology, 1999.

[8] Mintz M, Bills S, Snow R, and Jurafsky D, Distant Supervision for Relation Extraction Without Labeled Data, Joint Conference of 47th Annual Meeting of the ACL and 4th IJCNLP of the AFNLP, 2009.

[9] McCallum A, Freitag D, and Pereira F, Maximum Entropy Markov Models for Information Extraction and Segmentation, ICML 2000.

[10] Lafferty JD, McCallum A, and Pereira FCN, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, ICML 2001.

[11] Rabiner LR, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 77(2):257-285, 1989.

[12] Florian R, Hassan H, Ittycheriah A, Jing H, Kambhatla N, Luo X, Nicolov N, and Roukou S, A Statistical Model for Multilingual Entity Detection And Tracking, HLT/ NAACL 2004.

[13] Berger AL, Della Pietra S, and Della Pietra VJ, A Maximum Entropy Approach to Natural Language Processing, Computational Linguistics, 22(1): 39-71, 1996.

[14] Culotta A and McCallum A, Reducing Labeling Effort for Structured Prediction Tasks, AAAI 2005.

[15] Manning CD, and Schutze H, Foundations of Statistical Natural Language Processing, The MIT Press, 1999.

Address for correspondence

Jian Ni

IBM T. J. Watson Research Center

1101 Kitchawan Road, Yorktown Heights, NY 10598, USA

nij@us.ibm.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-40

Using EHRs and Machine Learning for Heart Failure Survival Analysis

Maryam Panahiazar[a], Vahid Taslimitehrani[b], Naveen Pereira[c], Jyotishman Pathak[a]

[a] *Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA*
[b] *Department of Computer Science and Engineering, Kno.e.sis Center, Wright State University, Dayton, OH, USA*
[c] *Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA*

Abstract

"Heart failure (HF) is a frequent health problem with high morbidity and mortality, increasing prevalence and escalating healthcare costs" [1]. By calculating a HF survival risk score based on patient-specific characteristics from Electronic Health Records (EHRs), we can identify high-risk patients and apply individualized treatment and healthy living choices to potentially reduce their mortality risk. The Seattle Heart Failure Model (SHFM) is one of the most popular models to calculate HF survival risk that uses multiple clinical variables to predict HF prognosis and also incorporates impact of HF therapy on patient outcomes. Although the SHFM has been validated across multiple cohorts [1-5], these studies were primarily done using clinical trials databases that do not reflect routine clinical care in the community. Further, the impact of contemporary therapeutic interventions, such as beta-blockers or defibrillators, was incorporated in SHFM by extrapolation from external trials. In this study, we assess the performance of SHFM using EHRs at Mayo Clinic, and sought to develop a risk prediction model using machine learning techiniques that applies routine clinical care data. Our results shows the models which were built using EHR data are more accurate (11% improvement in AUC) with the convenience of being more readily applicable in routine clinical care. Furthermore, we demonstrate that new predictive markers (such as co-morbidities) when incorporated into our models improve prognostic performance significantly (8% improvement in AUC).

Keywords:

Heart Failure; Survival Score; Electronic Health Records; Machine Learning.

Introduction

Heart failure (HF) is primarily caused by the inability of the heart to supply sufficient blood flow to the body. It has become one of the most deadly cardiovascular diseases in the 21st century [1]. Therefore, it is important to identify patients who are at a higher risk of mortality due to HF and assess the impact of HF therapy on their outcomes. Several studies have developed prognostic tools for HF, and one of the most commonly used tools is the Seattle Heart Failure Model (SHFM) [1]. SHFM was based on the PRAISE I clinical trial database and validated in five other cohorts.

While the derivation and validation of SHFM using such clinical trials databases provide a level of rigor in terms of data collected, they are typically limited to biased cohorts that tend to be homogenous. Further, in many cases, the environment in which the trials were conducted may not reflect the routine clinical care given to patients diagnosed with HF. Consequently, as EHRs become more ubiquitous and accessible for clinical research, it becomes imperative to investigate methods of predicting HF prognosis and the impact of HF therapy on important patient-related outcomes using EHRs–as opposed to data derived exclusively from clinical trials databases. The specific objectives of this study are to assess the performance of the SHFM using routinely collected EHR data in a community practice at Mayo Clinic, and to incorporate variables that were not part of the SHFM in our prognostic model (e.g., patient co-morbidities derived from the EHR) to assess improvement in the performance of survival analysis.

Our results suggest that (1) heart failure survival models built on EHRs are more accurate than the SHFM, (2) incorporating co-morbidities into the heart failure survival analysis prediction models improve the accuracy of our models, and (3) there are potential hidden interactions between diagnoses history of the patient, co-morbidities, and survival risk. We also build our models using multiple different machine learning algorithms and our results show that logistic regression and random forest return more accurate classifiers.

Background and Related Work

The SHFM was derived in a cohort of 1125 heart failure patients from the PRAISE I clinical trial with the use of a multivariate Cox model [1]. For variables such as medications (e.g., beta-blockers) and devices (e.g., defibrillators) that were not available in the derivation database, hazard ratios were estimated from published literature and "external" clinical trials. The model has been prospectively validated in 5 additional cohorts totaling 9942 heart failure patients and 17307 person-years of follow-up.

However, the SHFM has significant limitations for risk prediction in HF, particularly when used for routine clinical care. For instance, the hazard ratios for a subset of medications and devices variables in SHFM were estimated from prior published literature, and results from prior clinical trials may not be generalizable to a wider real-world population of HF patients. Limited patient-specific parameters (e.g., patient co-morbidities) have been used in the model to calculate survival score and can potentially lead to an improvement in the prediction of HF prognosis.

In addition to SHFM, several other risk prediction models have been developed including SHOCKED, Frankenstein, PACE Risk Score, and HFSS [1]. These have been validated in independent cohorts along with SHFM: "The Heart Failure Survival Score (HFSS) was validated in 8 cohorts (2240 patients), showing poor-to-modest discrimination (c-statistic, 0.56–0.79), being lower in more recent cohorts. The Seattle Heart Failure Model was validated in 14 cohorts (16,057 patients), describing poor-to-acceptable discrimination (0.63–0.81), remaining relatively stable over time. Both models re-

ported adequate calibration, although overestimating survival in specific populations. The other 3 models were validated in a cohort each, reporting poor-to-modest discrimination (0.66–0.74) [1-5].

Furthermore, there are also studies that applied machine-learning algorithms to study risk factors and predict patient outcomes in HF. For example, Dai et al. [6] used boosting and support vector machine (SVM) schemes to build models to predict heart failure around six months before the actual diagnosis. Their results show that SVM has poor performance. Similarly, Austin et al. [7] used regression tree, bagging, Random Forest, boosting, SVM and logistic regression to classify HF patients with preserved Ejection Fraction (EF) from those patients with reduced EF. They concluded that logistic regression returns the most accurate models.

Methods

From a cohort of 119,749 Mayo Clinic patients between 1993-2013 with research authorization to access EHR data, we identified 5044 patients with a diagnosis of HF after applying specific criteria and excluding number of patients due to incomplete data (13.3%). These criteria included:

- A confirmed diagnosis of HF based on the ICD-9-CM code (428.x).
- An ejection fraction (EF) measurement ≤50% within two months of HF diagnoses.
- No prior diagnosis of coronary artery disease, myocarditis, infiltrative cardiomyopathy, and severe valvular disease.

We divided the EHR-derived dataset randomly into training (N=1560 patients) and test (N=3484 patients) datasets. In consultation with Mayo Clinic cardiologists who routinely treat HF patients, we identified the following features (variables) extracted from EHRs to calculate the survival score:

- Demographic variables including age, sex, race, ethnicity and survival status.
- Laboratory results including cholesterol, sodium, hemoglobin, lymphocyte count, and EF measurements.
- Medications including Angiotensin Converting Enzyme (ACE) inhibitors, Angiotensin Receptor Blockers (ARBs), β-adrenoceptor antagonists (β-blockers), Statins, and Calcium Channel Blocker (CCB).
- 26 major chronic conditions (ICD-9 code) as co-morbidities as defined by the U.S. Department of Health and Human Services [8].

Table 1 represents the characteristics of the study cohort with including training and test splits. The patients were 94% white 48% female. The average range of EF was 36±10.3. In terms of co-morbidities most of the patients (81.06%) had hypertension followed by hyperlipidemia (64.3%), chronic kidney disease (55.83%) and diabetes (37.4%).

As discussed earlier, our primary goal in this study is to assess the performance of SHFM using EHR data and propose a HF survival risk prediction model by adding new variables (e.g., patient co-morbidities) derived using the EHR data to improve prediction accuracy and performance of the model. To accomplish this goal, we designed two scenarios: In scenario A, we used COX proportional regression model [9] to predict the risk of survival in HF patients in the one, two and five years after the diagnosis of HF. In scenario B, we excluded all patients who died within one year after the first diagnosis of HF, and then based on the remainder of patients who survived after one year of HF diagnosis, we developed a series of models using different classifiers to classify these two groups of patients. Since most of the well known classification algorithms are developed for binary classification, we repeated scenario B to classify patients who died within two years and five years after the HF diagnosis. The following classification algorithms were used: random forest [10], logistic regression [11,12], support vector regression [13], decision tree [14] and ada boost [15].

Table 1 – Patient Characteristics for HF Study Cohort

	Variables	Value
Demographic	Age (years)	78±10
	Sex (male)	52%
	Race (White)	94%
	Ethnicity (Not Hispanic or Latino)	84%
Laboratory	BMI	28.7±11.25
	Systolic Blood Pressure (mm/Hg)	120±25
	Ejection Fraction (EF)	36±10.3
	Hemoglobin (g/dL)	11.8±1.2
	Cholesterol (mg/dL)	144±35
	Uric Acid (g/dL)	7.1±2.5
	Sodium (mEq/L)	128±4.2
	Lymphocytes (x10(9)/L)	1.32±0.7
Medications	ACE inhibitors	55.7%
	Beta blockers	48.6%
	Angiotension Receptor Blockers	12.8%
	Calcium Channel Blockers	4.1%
	Statins	43.2%
	Diuretics	68.7%
	Allopurinol	18.5%
	Aldosterone Blockers	18.5%
Comorbidities	Hypothyroidism	21.2%
	Acute myocardial infarction	16.3%
	Alzheimers	11.9%
	Anemia	5 3.01%
	Asthma	10.72%
	Atrial fibrillation	48.56%
	Benign prostatic hyperplasia	9.5%
	Cataract	31.4%
	Chronic Kidney Disease	55.83%
	Pulmonary disease	30.4%
	Depression	25.5%
	Diabetes	37.4%
	Glaucoma	9.4%
	Hip/pelvic fracture	4.3%
	Hyperlipidemia	64.3%
	Hypertension	81.06%
	Ischemic heart disease	70.2%
	Osteoporosis	18.3%
	Rheumatoid Arthritis	39.2%
	Stroke	12.4%
	Breast cancer	2.2%
	Colorectal cancer	1.58%
	Prostate cancer	4.5%
	Lung cancer	2.45%
	Endometrial cancer	0.00%

To investigate the effect of variables extracted from the EHR data on the performance of our models, we designed two sets of predictor variables. In the first set called *baseline* (*BL*), we applied the same variables used in the SHFM. Since our EHR

derived dataset does not have information about patients' NYHA class, QRS duration, or device implantations (e.g., defibrillators), we did not include them in our models. In the next variable set called *extended (EX)*, we added the following predictor variables to the *BL* model: race, ethnicity, BMI, calcium channel blocker (CCB) and 26 different co-morbidities. Then we compared the performance of our *BL* and *EX* models with the SHFM.

Results

This section reports a systematic validation of our HF survival risk prediction model(s) in both scenarios A and B. As we mentioned earlier, to minimize the effect of overfitting and increase generalizability of our models, we separated our cohort randomly into training (N=1560 patients) and test (N=3484 patients) datasets. To validate models which are developed in scenario A (survival models), we designed two approaches. In the first approach (A1), we divided our predictor variables into two parts: variables that were common between both the SHFM model and our models (e.g., EF measurement), and variables used by just our model (e.g., patient co-morbidities). For variables that were common we used the hazard ratios defined by SHFM. For variables that were not used in SHFM we used the hazard ratios extracted from our analysis. Table 2 presents the performance of our model (A1) in the form of Area Under Curve (AUC). In the second approach (A2) we used the hazard ratio developed by our model to calculate the AUCs (also shown in Table 2).

Table 2 shows that accuracy drops on average by 9% in the AUC when using a mixture of hazard ratios from our model and SHFM. Although it is not surprising, one plausible reason might be that SHFM was developed using a clinical trials database, whereas our model has been developed using an EHR database in a real community practice. Consequently, combining hazard ratios derived from two different datasets can lead to a degradation in the accuracy of the prediction model.

Table 2 – Perfomance (Area Under Curve) of Scenario A (Survival Analysis) for both baseline and extended variable sets

	1 year		**2 years**		**5 years**	
	BL	EX	BL	EX	BL	EX
A1	0.71	0.80	0.68	0.75	0.66	0.71
A2	0.77	0.83	0.77	0.82	0.75	0.80

Figure 1 also represents the Receiver Operating Characteristics (ROC) of the models developed in scenario A on the BL variable set. From Table 2 and Figure 1, we can observe that the performance of our models drops when we want to predict the risk of heart failure two and five years after the first heart failure event. We hypothesize that the main reason of observing a drop in the accuracy is because the dataset for two and five year models are imbalanced and usually the classifiers used in this study do not work well on imbalanced datasets. Figure 2 shows the ROC curve of the scenario A models on EX variable set.

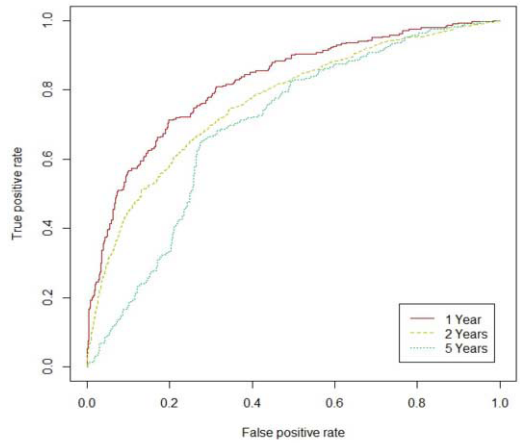


Figure 1 - ROC curve for Scenario A on baseline variable sets

As we discussed earlier, one of our goals in this study was to investigate the effect of adding more predictor variables to our models, and potentially improve model performance. To this end, we designed two variable sets called *BL* and *EX*. In the *BL* set, we have 16 predictor variables and in the *EX* set, we consider 45 variables to develop our models. Table 2 shows that the AUC for models developed using the *EX* variable set increased by 7.7%, 6.5% and 6.6% compared to the *BL* variable sets for 1-, 2- and 5-years models, respectively. In the next part of our results, we show the performance of the classification models (scenario B). In this scenario, we excluded all patients who died within 1-, 2- and 5-years after the first diagnosis of HF, and then developed models to classify them separately from patients who did survive after the HF diagnosis.

Table 3 – Perfomance of Senario B (Classification) for both baseline and extended variable sets

	1 year		2 years		5 years	
	BL	EX	BL	EX	BL	EX
Decision Tree	0.60	0.66	0.50	0.50	0.50	0.50
Random Forest	0.62	**0.80**	0.65	0.72	0.62	0.72
Ada Boost	0.59	0.74	0.66	0.71	0.61	0.68
SVM	0.56	0.46	0.61	0.52	0.55	0.38
Logistic Regression	**0.68**	**0.81**	0.7	0.74	0.61	0.73

Table 3 represents the AUC of different classifiers for both BL and EX variable sets. Much like scenario A, we make the same observation in scenario B where inclusion of additional variables derived from the EHR to the models significantly improves classifier performance.

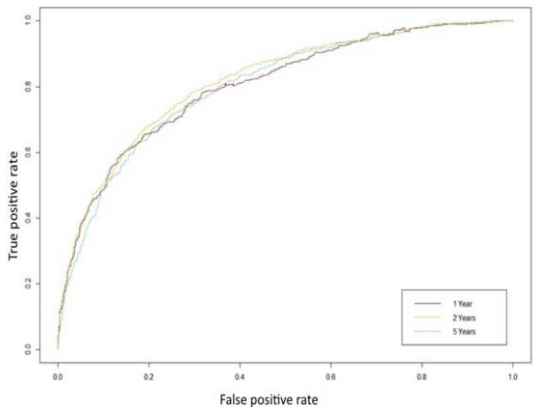
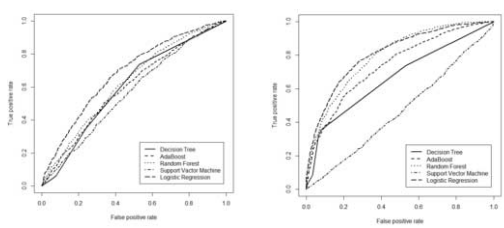
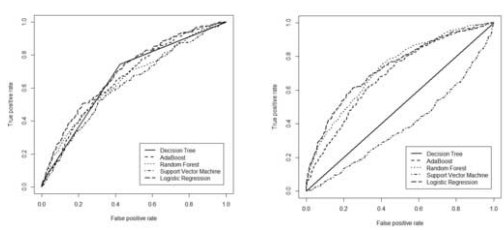


Figure 2 - ROC curve for Scenario A on exdended variable sets

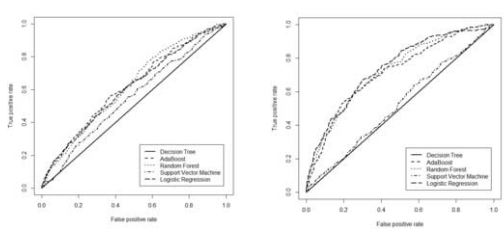
Figure 3 also represents the ROC of the models developed in scenario B on the BL and EX variable sets for one, two, and five years for all classifiers shown in Table 2, including Decision Tree, Random Forest, AdaBoost, SVM, and Logistic Regression. There are a number of reasons why SVM may have been less accurate in developing this prediction model. First, SVM is not an appropriate method for handling both continuous and categorical variables in the same model. Second, our data suggest that SVM may be more strongly affected by classification imbalance in the data than either Boosting or logistic regression.



BL model, 1 year *EX model, 1 year*



BL model, 2 years *EX model, 2 years*



BL model, 5 years *EX model, 5 years*

Figure 3 - ROC curve for Scenario B on BL & EX variable sets(x=False positive rate, y=True positive rate)

Discussion

In this study we explored how to improve SHFM by considering routine clinical care data. Since models that are built based on EHRs are more accurate (11% improvement in AUC) and are applicable in standard routine care, it is imperative to leverage EHR data for survival analysis and prediction modeling in HF and other chronic conditions. We also showed that incorporating new predictive markers (co-morbidities) in our models improved the performance significantly (8% improvement in AUC) and gives us insights about the pathophysiology of HF. Another highlight of this study is calculating the Hazard Ratio (HR) based on real-world EHR data, whereas other studies, including SHFM, have used the HR from literature and extrapolation of results from clinical trials which may not reflect routine care for HF patients [1,5]. Finally, we observe that there are potential hidden interactions between diagnoses, history of the patient, co-morbidities, and survival risk that warrant further research. Note that since we calculate the model output based on individual patient chracteristics, it is plausible to incorporate the output derived from these predictive models within EHRs and facilitate clinical decision making for managing HF patients with better treatment options–an area for future research. In summary, our results suggest that heart failure survival models built on EHRs are more accurate, and incorporating co-morbidities into the HF models significantly improves the accuracy of our models.

Conclusion

In this study we assessed the performance of SHFM using Mayo Clinic's EHR dataset. Our results demonstrate an improvement in accuracy as compared to the standard SHFM and also suggest the ready applicability of our model to standard clincal care in the community. We also incorporated additional predictor variables that included 26 co-morbidities into our models that lead to further improvement in the prognostic predictive accuracy. Finally, we built a heart failure risk prediction model using a series of machine learning techniques and observed that logistic regression and random forest return more accurate models compared to other classifiers.

Acknowledgments

This project was funded in part by support from AHRQ (R01 HS023077).

References

[1] Alba AC, Agoritsas T, Jankowski M, et al. Risk Prediction Models for Mortality in Ambulatory Patients with Heart Failure: A Systematic Review. Circulation. Heart failure, vol. 6, no. 5, pp. 881–9, Sep. 2013.

[2] Cabassi A, Champlain J, Maggiore U, et al. Prealbumin. Improves Death Risk Prediction of BNP-added Seattle Heart Failure Model: Results from a Pilot Study in Elderly Chronic Heart Failure Patients. International Journal of Cardiology, vol. 168, no. 4, pp. 3334–9, Oct. 2013.

[3] Hussain S, Kayani AM, Munir R, et al. Validation of the Seattle Heart Failure Model (SHFM) in Heart Failure Population. Journal of the College of Physicians and Surgeons–Pakistan: JCPSP, vol. 24, no. 3, pp. 153–6, Mar. 2014.

[4] Levy WC, Mozaffarian D, Linker DT, et al. The Seattle

Heart Failure Model: Prediction of Survival in Heart Failure. Circulation, vol. 113, no. 11, pp. 1424–33, Mar. 2006.

[5] Prasad H, Sra J, Levy AC, and Stapleton DD. Influence of Predictive Modeling in Implementing Optimal Heart Failure Therapy. The American Journal of the Medical Sciences, vol. 341, no. 3, pp. 185–90, Mar. 2011.

[6] Dai W, Brisimi T S, Adams W G, Mela T, Saligrama V, and Paschalidis L C. Prediction of hospitalization due to heart diseases by supervised learning methods. International Journal of Medical Informatics. 2014.

[7] Austin P C, Tu J V, Ho J E, Levy D, and Lee D S. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. Journal of Clinical Epidemiology, 66.4: 398-407, 2013.

[8] U.S. Department of Health & Human Services http://www.hhs.gov/ash/initiatives/mcc/

[9] Fox J. Cox Proportional-Hazards Regression for Survival Data, Appendix to An R and S-PLUS Companion to Applied Regression, February, 2002.

[10] Liaw A, and Wiener M, Classification and Regression by Random Forest. R news 2.3: 18-22, 2002.

[11]Hosmer D, Lemeshow S, and Sturdivant RX. Introduction to the logistic regression model. John Wiley & Sons, Inc., 2000.

[12]Taslimitehrani V, Dong G. A new clinical prediction method using contrast pattern aided logistic regression with application on traumatic brain injury. In IEEE International Conference on BioInformatics and BioEngineering (BIBE), Nov 2014. Best Student Paper Award.

[13]Furey TS, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16.10: 906-914, 2000.

[14]Safavian SR, and Landgrebe D. A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 21.3 1991, 660-674, 2000.

[15]Collins M, Schapire RE, and Singer Y. Logistic regression, AdaBoost and Bregman distances. Machine Learning 48.1-3, 253-285, 2002.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-45

Towards the Implementation of an openEHR-based Open Source EHR Platform (a vision paper)

Pablo Pazos Gutiérrez[a]

[a] *openEHR en Español, Asociación Chilena de Informática en Salud, CaboLabs*

Abstract

Healthcare Information Systems are a big business. Currently there is an explosion of EHR/EMR products available on the market, and the best tools are really expensive. Many developing countries and healthcare providers cannot access such tools, and for those who can, there is not a clear strategy for the evolution, scaling, and cost of these electronic health products. The lack of standard-based implementations conduct to the creation of isolated information silos that cannot be exploited (i.e. shared between providers to promote a holistic view of each patient's medical history).

This paper exposes the main elements behind a Standard-based Open Source EHR Platform that is future-proof and allows to evolve and scale with minimal cost. The proposed EHR Architecture is based on openEHR specifications, adding elements emerged from research and development experiences, leading to a design that can be implemented in any modern technology. Different implementations will be interoperable by design. This Platform will leverage contexts of scarce resources, reusing clinical knowledge, a common set of software components and services.

Keywords:

Electronic Health Records; Open Source Software; openEHR; Service Oriented Architecture; Data Collection.

Introduction

In this vision paper we present the basic components of an Open EHR Platform that will be used in different scenarios as a shared EHR core for country wide, federated, hospital and clinic EHRs. The term "EHR core" will be use to reference an implementation of the Platform using a specific technology stack.

The main motivation behind this EHR Platform is to create an open alternative to the closed/proprietary solutions currently offered by EHR vendors. But also to create an EHR Platform that is based on open standards and good design practices, searching for a generic, reusable, scalable, extensible, web-based, knowledge-driven, and future-proof solution.

This Platform is “open” in three ways:

1. open specification accessible for anyone.
2. open source tools and components.
3. free to use, adapt and extend.

This will help developers to create new tools and services without the hassle of dealing with e-Health standards and clinical knowledge (time consuming and costly tasks of any EHR development), and enable developments in contexts of scarce human & financial resources (small software companies, dev teams in hospitals, etc.).

In the next sections the following topics will be presented: design principles, architecture, components, services, and ideas pertaining to Platform development and use by client applications. Developers using services provided by the EHR Platform, for shared data storage, querying and Clinical Decision Support (CDS), will create these applications.

Current Status

Although this paper is about the vision and conception of a new kind of EHR systems, we and other colleagues [1][2] have been working in this area, including research, development and training, mainly to understand the problem, delineate a good solution, and validate it against functional prototypes of the EHR Platform. The outstanding component we created is the EHRServer [3]. It provides services of clinical data storage and querying for clinical information, and this system can actually be deployed and used. For Knowledge Management, the Clinical Knowledge Manager [4] of the openEHR Foundation is being used. There are also some proof-of-concept developments around a Demographic Server and a Rule Engine that will soon be released.

Methods

This paper presents the core components of the proposed EHR Platform, which is an extension/specialization of the openEHR EHR Computing Platform [5], defined by the openEHR specifications [6]. The extensions are based on research & development, project development, integration of EHR, HIS and other clinical information systems, and preparing training materials and case studies.

Design Principles of the Open EHR Platform

- Provide a minimum/common set of generic services
  - Reuse: must support a wide range of client applications with the same set of components and services, leveraging available resources.
  - Pareto principle: the platform will provide the 20% of the functionalities of a full blown EHR system to support 80% of the requirements for any EHR system, keeping the Platform small and manageable. The rest of the functionalities should be added by developing client applications, and those might be customized for each application (i.e. won’t suit as core functionalities).
  - Scope: The services will be focused on data storage and querying, permitting data interpretation and manipulation to client applications.
  - Standardization: the EHR Platform will implement open standards to reach interoperability, so developers do not need to do that work.

- Modifiability and Extensibility: the metadata-based configuration of the Platform will allow new data structures, data queries and rules to be added without modifying the database structure or source code.
- Clinical data consolidation:
  - The storage services of the Platform collect and consolidate clinical information from different client applications into a loosely centralized platform, thus facilitating a unique EHR, per patient, that is complete and ubiquitously accessible.
  - Consolidated data will be on a Cloud-ready Platform using a distributed approach, avoiding monolithic systems and one-point-of-failure.
- Secure:
  - Clinical and demographic/identification data are physically separated, allowing secondary use of anonymized clinical data.
  - All communications between client applications and the EHR Platform will be encrypted. On each communication, the client will be authenticated and permissions will be verified. Client apps should comply with security rules, and good security practices will be encouraged.
  - Modifications to the data in the EHR will be done only by authenticated and authorized client applications, and those should ensure correct authentication and authorization for their users (this item should be included in the Platform rules of use).
  - No data is actually modified or deleted; amendments, corrections or deletions all generate new versions of the data.

Knowledge Management (KM)

The KM process is grounded on four main artifacts: 1) clinical record structures, 2) terminology, 3) rules, and 4) processes. Those are defined as computable expressions that can be created and updated, using modeling tools [7], by Domain Experts (physicians, nurses, and allied health professionals). These experts, who are closer to the end users of the EHR, will play the role of Clinical Knowledge Modelers that are in responsible for defining the structure and behavior of the EHR, and how it will evolve over time.

The openEHR specifications [6] define two kinds of artifacts: archetypes and templates. Archetypes are units of content (data structures, constraints and terminology) that specify to one health concept, in a processable and self-contained artifact, that has global validity (e.g. Blood Pressure, Problem/Diagnosis, Clinical Evolution, Glasgow Coma Scale, Body Mass Index, etc.). Templates defined by specific clinical documents, valid in local contexts, refer to many Archetypes, and can use all the Archetype structure or just part of it.

Open EHR Platform Architecture

The Open EHR Platform Architecture is constituted by components with well-defined responsibilities, and provides services to other and client applications. This is based on the openEHR architecture [8].

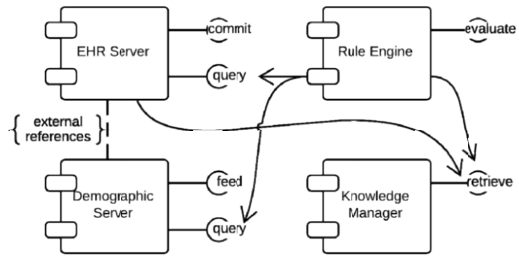


Figure 1 - Open EHR Platform Architecture

Four main components are defined to manage information and provide services for: 1) clinical data (EHR Server), 2) demographic and identification data (Demographic Server), 3) rules to enable CDS (Rule Engine), and 4) Clinical Knowledge (Knowledge Manager). The latter is where all of the metadata, authored by Domain Experts, is stored and made available for the rest of the components: clinical document and administrative data structures, terminologies, rules and workflows.

Due to the narrowed scope of the Platform, a User Interface (UI) for the end-user ought to be provided by client applications developed over the EHR platform. This is why the term "EHR platform" rather than "EHR system" may better suit this work.

Services

Each sub-system of the EHR Platform architecture will provide a set of services related to their responsibilities. Below we describe the major services and internal components of each sub-system.

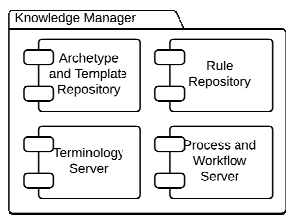


Figure 2 - Knowledge Manager Component

The KM sub-system will handle the artifacts created by the KM process. These artifacts are the heart of the whole EHR Platform, if some of those change or new artifacts are added, the data and behavior of the Platform will change, giving the Platform the "future-proof" label. There are four main components:

- Archetype and Template Repository: stores artifacts that define the data structure and low-level constraints of each clinical document in the EHR.
- Terminology Server: provides services around terminologies, classifications, thesauri, dictionaries, and mappings/relationships between them. Archetypes and Templates will use these terminologies.
- Rule Repository: manages rules that will be executed by the Rule Engine.

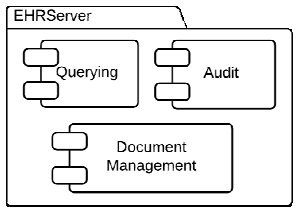


Figure 3 - EHR Server Components

The **EHR Server** is responsible for the safekeeping of clinical records and providing services to change the EHR and query for clinical data.

Document Management: Provides services to validate, create and update clinical information in the EHR of each patient, allowing the management of the internal organization of the clinical records in a directory structure (e.g. by episode or health problem).

Querying: Enables client applications such as EMRs and mHealth apps to access clinical information from the EHR. Queries are specified in an abstract declarative way, based on structures defined by openEHR archetypes [9]. Queries do not depend on specific database technologies, so there is the need to implement a crosswalk between EHR Server queries and database query language such as SQL or XQuery. In this component, queries will be created by Domain Experts and end-users [10], exported & imported and executed against a specific DB technology. So EHR queries can be shared between different implementations of the EHR Server. The results of a query will be returned in standard open formats such as XML and JSON, and organized in different ways: full clinical documents or data points, grouped by data type or by clinical document type, contextualized to time points or intervals.

Audit: Enables the management, evaluation and control of changes to the EHR of a patient, allows the detection of problems such as wrongly organized, deleted or modified clinical information.

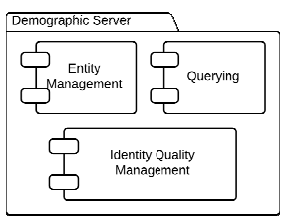


Figure 4 - Demographic Server Components

The Demographic Server is responsible for the management of varying entities such as persons, organizations, groups, devices, systems, their relationships and roles. All of these entities will be referenced from the EHR records, and all referenced entities should be stored in the Demographic Server. This sub-system will act as a Master Patient Index and Human Resource Index (but will not include information about contracts, payments, etc. that will be handled by third party applications). The Identity Quality Management includes functionalities to help detect errored, corrupted and duplicated entities in the demographic repository, and manage the repository audit to evaluate and control changes made by external applications.

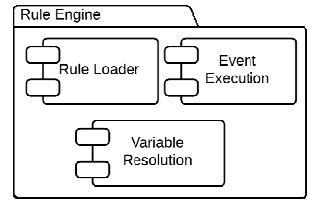


Figure 5 - Rule Engine Components

Rules are declarative units of logic that are evaluated against data, that execute actions based on conditions, and can return values. The rule execution will be triggered by events such as a user clicking a button or selecting a value from a list (e.g. a medication to create a prescription) or system events such as batch jobs executed at certain times. The data to evaluate the rules will come from the EHR Server (clinical data) or the Demographic Server (age range, sex, geographic location), and can be input data (provided by the triggered event), or rule variables (resolved when the rule is executed, e.g. querying the EHR for data). There is a proposal from the community to express rules and guideline-based workflows using a similar syntax to the one used in archetypes [11], this is called the Guideline Definition Language (GDL) [12].

The **Rule Engine** is in charge of loading rules from the Knowledge Manager, Resolve Referenced Values (e.g. querying the EHR Server), and Rule Execution that controls the rule logic and return values.

Customization and Extension: a better way

Since our goal is to design a low cost environment to create EHR systems, the Platform we envision will have an extreme level of modifiability, and this is a design requirement. To reach this objective we need to forget some of the traditional practices currently used to develop software for healthcare. In the traditional way we might be able to create an EHR, but we might not be able to maintain it. So, considering that the costlier and longer phase in software development is maintenance and evolution [13], and EHRs are long-term projects, thus for really effective EHR Systems we need to cut the maintainability cost to the minimum.

The Platform architectural design and the use of metadata to describe the system data structures (archetypes, templates, terminology) and behavior (rules, processes) allows a huge level of modifiability to adapt the EHR to new requirements, while maintaining the stability of the core EHR system. This is achieved because there is no need to modify source code or database schemas to make changes to the Platform, so no new bugs are introduced at the code level. To implement this, a mix of open standards, a Service Oriented Architecture (SOA), current technologies and techniques (Model-View-Controller, Object-Relational Mapping and dynamic languages) should also be applied.

Making changes and adding new features for end-users will be done over client applications, e.g. improving EMR UI or adding a feature, such as tagging patient records, so the core of the system, the EHR Platform, is not affected by these changes.

For example, on the EHR Server, extensions will be added in three main forms:

1. New archetypes and templates will define the structure of new clinical records,
2. New queries will define new ways of accessing the data stored in the EHR,

3. New rules will add behavior, to execute actions under certain conditions, using the data from the EHR.

Deployment

The proposed platform can be offered by a "Software as a Service" (SaaS) model, or software vendors and medical institutions can create their own local deployments. Either way, the platform is focused on providing high availability and data/metadata backup, two basic requirements for any shared EHR system. To provide this, some redundancy is needed; therefore more than one instance of each sub-system should be deployed.

The platform is designed to support this by providing synchronization services between different instances of the same sub-system. E.g. two instances of EHR Server will synchronize data when one instance receives a commit of a clinical document, so queries can be executed in both instances, also providing better performance. This also works when scalability is needed. When the EHR Platform needs to support more applications and users, store and retrieve more data, etc., the service level of the platform should neither deteriorate nor disturb other users. Adding more instances of the Platform sub-systems will let it to scale horizontally without huge investments.

Thinking about long-term viability of a project of this kind, offering SaaS as a hosted solution for software development companies can help to fund the EHR Platform maintenance and evolution. This strategy is implemented by a lot of Open Source projects, accompanying that with paid extensions and value-added services.

Developing Applications

The core set of services provided by the Open EHR Platform will enable developers to focus on creating client applications without starting from scratch, and solve the same problems again and again. Moreover, they get standardized data structures using communication protocols and syntaxes they are familiarized with like REST/SOAP, JSON/XML, without the hassle of implementing the standards themselves, so they create out-of-the-box interoperable applications.

Client applications can be classified into three categories: data input apps, data display apps and mixed. Vital Signs monitors and study result reporting are examples of "data input" applications, because they function to send/commit data to the EHR. Reporting tools are an example of "data display" applications that mainly will query the EHR. EMRs [14] are examples of mixed input/output. It is worth mentioning that when a deployed platform supports several applications, all the clinical documents committed to the EHR by the "data input" apps will be available to the "data display" apps with permission to query the EHR Platform, enabling the implementation of a truly shared EHR.

To help developers adopt the EHR Platform and the methodology to create client applications proposed in this paper, we visualize the creation of helper tools such as Software Development Kits, libraries for different programming languages and learning materials. Some work around this area has already been done. The first is a framework to develop EHR systems called EHRGen [15]. We have also worked in a multi-level methodology and a set of tools to automatically generate User Interfaces for Clinical Information Systems over different technologies (HTML5, Java, .Net), from openEHR Archetypes and Templates [16].

As an example, using the services provided by the EHR Platform, client applications can implement features like: browsing through the patient's medical history, generate clinical and public health reports about population's health with different aggregations (by age range, sex, geographic location, etc.), calculate for different indicators (e.g. number of births per year), sending notifications under certain circumstances to the patient, family, health care team, management team or healthcare authorities (e.g. execution of rules that notify when new test results are available, or when a diagnosis of a certain decease was added to the EHR of a patient), Clinical Decision Support features like alerts, recommendations, reminders, reference material, etc. (rules are executed, and based on context data, a correspondent result is returned, e.g. give an alert if the patient is allergic to the medications that is about to be prescribed).

Validation of the proposed architecture

The validity of the proposed architecture can, and should, be evaluated against standards and specifications, first against international ones, to satisfy generic requirements, methodologies, needs and rules, and then against local compliances that define more specific requirements. Since the architecture is the most generic and abstract definition of a complex system, it should not be evaluated against requirements related to low level designs, with a bigger amount and more specific requirements), and the use of specific technologies. After the architecture is adapted to a low level designs and mapped to a specific technology stack (this is called Implementation Technology Specification, ITS for short), those more specific artifacts can be evaluated to more specific requirements and rules.

Considering the aforementioned ideas, it would be useful to define some guidelines about how to validate the proposed architecture. The actual validation would be a good area for further investigation. Presented below is a list of varied standards and specifications that should (*) or might (+) be used as validation guidelines. In some cases just part of the standard would apply because it is too specific and the validation of the architecture should be done in a very broad, generic and abstract way, because that is the nature of a Software Architecture. After these, local standards and specifications might also apply.

Introduction, Terminology, Requirements and Architecture:

(*) ISO 18308: Requirements for an electronic health record architecture.

(*) openEHR Specifications [6]

(+) IHE IT Infrastructure Profiles [17]

(+) ISO 20514: Electronic health record -- Definition, scope and context

(+) ISO 14292: Personal health records -- Definition, scope and context

(*) ISO 14639: Capacity-based eHealth architecture roadmap -- Part 1: Overview of national eHealth initiatives

(+) ISO 14265: Classification of purposes for processing personal health information

(+) ISO 21298: Health informatics -- Functional and structural roles

(+) ISO 22790: Health informatics -- Functional characteristics of prescriber support systems

Healthcare record definition and sharing:

(*) ISO 27790: Document registry framework

(*) ISO 13128: Clinical document registry federation

(+) ISO/HL7 27932: Data Exchange Standards -- HL7 Clinical Document Architecture, Release 2

(+) ISO 12773-1: Business requirements for health summary records -- Part 1: Requirements

(+) ISO 12773-2: Business requirements for health summary records -- Part 2: Environmental scan

(+) ISO 13119: Clinical knowledge resources -- Metadata

Healthcare Records and Healthcare Data Communication:

(*) ISO 13606-1: Electronic health record communication -- Part 1: Reference model

(*) ISO 13606-2: Electronic health record communication -- Part 2: Archetype interchange specification

(*) ISO 13606-3: Electronic health record communication -- Part 3: Reference archetypes and term lists

(*) ISO 13606-4: Electronic health record communication -- Part 4: Security

(*) ISO 13606-5: Electronic health record communication -- Part 5: Interface specification

Conclusion

We are confident that creating a truly Open EHR Platform will enable e-Health developments and long term EHR projects, that are impossible today due the lack of resources. The proof-of-concept prototypes already developed are attracting attention from the academy and the industry, and we are very close to the delivery of EHR Platform sub-systems in a production-ready state. As time passes, we are also designing and developing tools that will benefit software developers in the creation of openEHR-based applications that will be compatible with the Platform. Such applications could produce a highly scalable backend for shared EHR systems, thus allowing their applications to share a common/stable set of services and easily evolve through time.

References

[1] Koray Atalag, et al., Assessment of Software Maintainability of openEHR Based Health Information Systems - A Case Study in Endoscopy, eJHI 2012; Vol7(1):e5

[2] Koray Atalag, et al., A Standards-based Approach to Development of Clinical Registries - Initial Lessons Learnt from the Gestational Diabetes Registry, HiNZ 2014.

[3] EHRServer: open source, openEHR based, Shared EHR System project. https://github.com/ppazos/cabolabs-ehrserver (accessed on Dec 22nd 2014)

[4] Clinical Knowledge Manager by Ocean Informatics http://ckm.openehr.org/ (accessed on Dev 22nd 2014)

[5] What's openEHR? http://www.openehr.org/what_is_openehr (accessed on Dec 22nd 2014)

[6] openEHR Specifications http://www.openehr.org/programs/specification/releases/1.0.2 (accessed on Dec 22nd 2014)

[7] openEHR Modeling Tools http://openehr.org/downloads/modellingtools

[8] Beale, T., Heard, S., openEHR Architecture Overview, openEHR Foundation, 2008. http://openehr.org/releases/1.0.2/architecture/overview.pdf

[9] Chunlan Ma, Heath Frankel, Thomas Beale, Sam Heard, EHR Query Language (EQL) – A Query Language for Archetype-Based Health Records, MEDINFO 2007, pp.397-401 (2007)

[10]Shelly Sachdeva, Subhash Bhalla, Visual Query Language for Archertype-Based Electronic Health Records Databases, Journal of Information Processing, Vol.20 No.2, pp.438-450, 2012.

[11]Thomas Beale, Sam Heard, Archetype Definition Language, openEHR Foundation 2008. http://www.openehr.org/releases/1.0.2/architecture/am/adl.pdf (accessed on Dec 22nd 2014)

[12]Rong Chen, Iago Corbal, Expressing and Sharing Clinical Decision Support Rules Using openEHR Archetypes, MEDINFO 2013.

[13]Schach, R. (1999), Software Engineering, Fourth Edition.

[14]Electronic Medical Record definition http://www.healthit.gov/providers-professionals/electronic-medical-records-emr (accessed on Dec 22nd 2014)

[15]Pablo Pazos, EHRGen: Generador de Sistemas Normalizados de Historia Clínica Electrónica Basados en openEHR, Congreso Argentino de Informática y Salud 2012.

[16]Arianne Palau, Laura Cuadrado, Pablo Pazos, Generación automática de interfaces de usuario para sistemas de información clínicos basados en una metodología multi-nivel, INFOLAC 2014.

[17]IHE IT Infrastructure Profiles http://wiki.ihe.net/index.php?title=Profiles#IHE_IT_Infrastructure_Profiles

Address for correspondence

Pablo Pazos Gutiérrez
Clinical Informatics Consultant
pablo.pazos@cabolabs.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-50

An Ecosystem of Intelligent ICT Tools for Speech-Language Therapy Based on a Formal Knowledge Model

Vladimir Robles-Bykbaev[a,b], Martín López-Nores[b], José Pazos-Arias[b], Diego Quisi-Peralta[a], Jorge García-Duque[b]

[a] GI-IATa, CIDII, Universidad Politécnica Salesiana, Cuenca, Ecuador
[b] AtlantTIC Research Center for Information and Communication Technologies, Department of Telematics Engineering, University of Vigo, Vigo, Spain

Abstract

The language and communication constitute the development mainstays of several intellectual and cognitive skills in humans. However, there are millions of people around the world who suffer from several disabilities and disorders related with language and communication, while most of the countries present a lack of corresponding services related with health care and rehabilitation. On these grounds, we are working to develop an ecosystem of intelligent ICT tools to support speech and language pathologists, doctors, students, patients and their relatives. This ecosystem has several layers and components, integrating Electronic Health Records management, standardized vocabularies, a knowledge database, an ontology of concepts from the speech-language domain, and an expert system. We discuss the advantages of such an approach through experiments carried out in several institutions assisting children with a wide spectrum of disabilities.

Keywords:
Electronic Health Records; Rehabilitation of Speech and Language Disorders; Medical Informatics.

Introduction

The language acquisition process is closely related with intellectual, psychological and emotional development of children. In the same line, speech and language serve as cornerstones for human cognition and constitute tools to interact with the fellows of the social environment. Commonly, during the first 4 years of life, children make significant progress in learning language. However, in some cases there arise delays or differences in patterns of language acquisition, aspects that constitute warnings of developmental problems [1]. These delays or differences can be related with a wide spectrum of circumstances and can appear in different stages of life. In the case of children, some disabilities are related with the following periods [2]: prenatal (chromosomal anomalies and genetic metabolic and neurologic disorders, drug and toxin exposure, congenital infections, ...) perinatal (complications related to prematurity, periventricular leukomalacia, breech or high forceps delivery, multiple births, placenta previa, preeclampsia, ...), or postnatal (undernutrition and environmental deprivation, viral and bacterial encephalitides, poisoning, ...). On the other hand, the speech, language and swallowing disorders in acquired neurological conditions arise as a result of several conditions, such as traumatic brain injury, stroke, progressive neurological conditions (Parkinson's disease, multiple sclerosis, ...), dementia, head and neck cancer, and palliative care [3].

In the same line, in a world where 1 billion people live with some form of disability, the most recent estimates about People with Communication Disabilities (PWCD) shown a complex picture. For example, nowadays 60 million persons in the world live with disabling hearing loss (5.3% of the world's population) and 15 million suffer from stutter, and there exists a lack of rehabilitation services and healthcare related to Speech-Language Disorders (SLD) in most of the developed countries [4]. The situation is even more complex in developing countries due to lack of personnel, resources, and adequate services.

Over the last years, there have been several approaches to apply Information and Communication Technologies (ICT) to support Speech and Language Therapy (SLT). However, the existing approaches to provide support tools for SLT deal with very specific problems. The aphasia treatment field [5] proposes a computer gesture therapy tool (GeST) with the aim to improve the gesture production and/or spoken naming, whereas [6] it investigates the feasibility to incorporate the Semantic Feature Analysis technique in a mobile-web paired application in domestic and clinical settings. Some speech disorders like dysarthria have motivated the application of Automatic Speech Recognition (ASR) techniques for diagnosis tasks [7]. On the other hand, the application of Auditory Verbal Therapy (AVT) to develop several speech abilities in children with cochlear implants has been successfully applied through videogames [8] and multimedia programs [9].

Nevertheless, nowadays some of the most important issues of health care field are related to maintenance, sharing and portability of clinical data, a perspective based on specific domain knowledge, and the adaptability and standardized vocabularies. Accordingly, the aforementioned approaches rely on ad hoc solutions that use particular vocabularies, languages and own data structures. Only a few recent studies have used conceptualizations and classifications (e.g. ontologies) to support the operation of an expert system aimed at the initial diagnosis of language disorders [10].

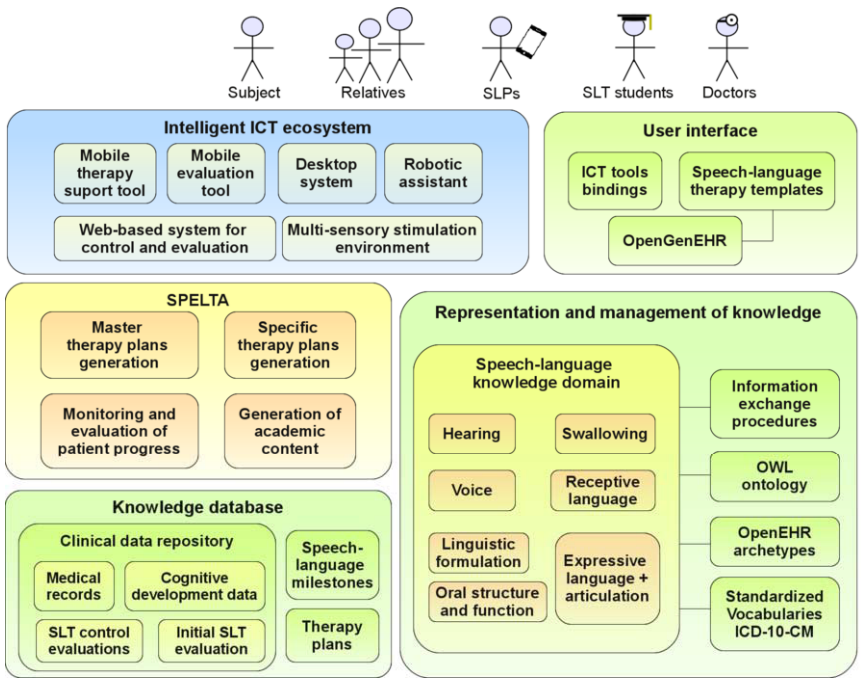


Figure 1– The high-level design and main components of our SLT ecosystem

Methods

SLT is a healthcare area that involves the active participation of different actors, such as doctors, Speech-Language Pathologists (SLP), SLT students, and patients/subjects and their relatives. Therefore, in this section, we present a complete backing ecosystem for SLT that considers these actors from a viewpoint that arises in a formal knowledge model. Likewise, this model is capable to provide an integrative perspective for ICT tools, specific SLT domain knowledge, standardized vocabularies, expert systems, and electronic health records management. Figure 1 shows a high-level model of the architecture of the proposed ecosystem for supporting SLT. This architecture allows building an integrative multi-layer framework for several therapy-related activities, users, knowledge management and portability, intelligent systems, and ICT tools. Some of the most relevant elements are described below.

The expert system layer (SPELTA)

The Speech and Language Therapy Assistant (SPELTA) is an expert system that provides the foundations to develop inference mechanisms for recommender and decision-support systems to assist in the preparation of therapy plans, the evaluation of exercise results, the generation of case studies, etc. Some of the most outstanding functionalities of SPELTA are the following:

- *Master therapy plans generation*: these kind of plans contain the general guidelines and activities that should be conducted with patients over a period of 6 months. Their aim is to develop some specific skills related to speech and language. In order to automatically generate a new master therapy plan, SPELTA uses a custom version of *Partition Around Medoids* (PAM) algorithm and KNN criterion (see more details in [11]). The PAM algorithm generates several clusters organized in two abstraction levels. The first level groups patients according to some elements of their profiles (cognitive developmental age, chronological age), medical record (diagnosis, related disorders/disabilities, and general medical condition), and the initial diagnostic of speech-language. In the second level several subclusters are created inside the general clusters, using the fine-grained evaluation of patient's speech-language skills.
- *Specific therapy plans generation*: these plans contain specific exercises and therapy activities that should be conducted with patients according to a weekly schedule. A set of specific therapy plans belongs to a master plan, and currently we are developing a semantic network using ontologies to represent the different relations between disabilities, disorders, and speech-language milestones and skills according to chronological and developmental age.
- *Generation of academic content:* the expert system is able to generate educational content for phonoaudiology students. These contents include tests about real and generated cases, training resources for decision making about therapy strategies, standardized vocabularies entries, etc.
- *Monitoring and evaluation of patient progress:* this feature of SPELTA allows to automatically analyze the patient progress and generate alerts for SLP when the master and specific plans are not effective.

Knowledge database

The knowledge database stores several information structures to support the data analysis tasks and inference process performed by the expert system layer:

- *Clinical Data Repository* (CDR): consolidates data from several clinical resources (medical records, cognitive development data, SLT control evaluations, initial SLT evaluation) and presents a unified view of patients. Our CDR uses standardized vocabularies from the American Speech-Language-Hearing Association [12].
- *Therapy plans*: these are data structures to represent either long term activities and exercises (master plans) and sets of activities scheduled according to patient's skills and therapy sessions (specific plans).
- *Speech-language milestones*: defines a set of skills according to chronological and developmental age (e.g. formal expressions of facts like "*from 1 to 2 years of normal development, a child must to be able to acquire new words on a regular basis, know a few parts of the body and point to them when asked*").

Representation and management of knowledge

In order to provide a comprehensive support for all therapy process, the model considers seven areas of speech and language: hearing, swallowing, voice, receptive language, linguistic formulation, oral structure and function (oral peripheral mechanism), and expressive language and articulation. The main processes, tests, evaluations, protocols, and therapy activities are modelled using ontologies and are implemented through OpenEHR archetypes. Figure 2 shows a screenshot of an archetype to conduct the Pure Tone Audiometry (PTA) test. As we can see, through this archetype it is possible to assess the patient's response to sound stimulation and voice commands, and determine if he/she is able to localize sound sources without visual stimuli.

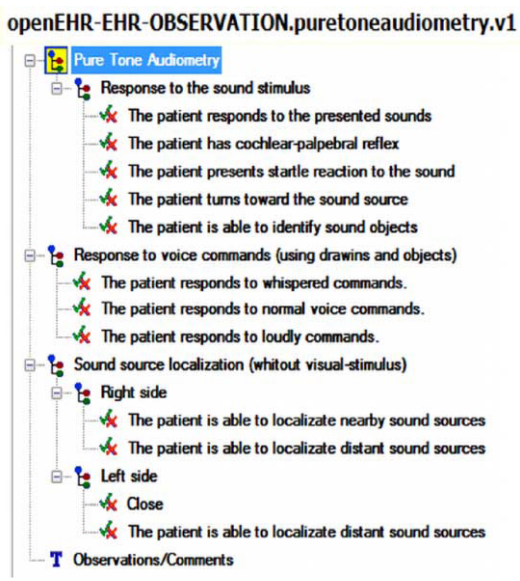


Figure 2– Partial view of the archetype that represents the Pure Tone Audiometry test

Intelligent ICT ecosystem

This layer incorporates a complete set of novel ICTs designed to support most of the SLT stages, and can easily add new tools using the bindings of the user's interface layer. Some of the most relevant elements are the following:

- *Multi-sensory stimulation ecosystem*: these kind of tools serve to support relaxation activities, early stimulation and several rehabilitation processes, for both children and adults, as was proved in [13] and [14].
- *Robotic assistant:* is a novel backing tool for therapy sessions, educational and ludic activities, and remote monitoring tasks. The robot has two main elements, an electronic displacement platform and a mobile device (Android-based tablet or smartphone). Using the mobile device the robot can tell stories, conduct exercises for motor rehabilitation (using a red glove to control a virtual hand), recognize voice commands, and interact with patients through different kinds of exercises based on multimedia resources (color discrimination, temporo-spatial notions, ...).
- *Support tools based on ICT:* through these tools the SLP is able to use desktop and web systems (Speech-language Maturation Assessment Tool, SPELMAT), and mobile devices to perform therapy exercises, assessment tasks, collection of information related with patients, registering the results of therapy sessions, report generation, and sharing data with other SLP. For example, Figure 3 (a) shows a screen capture of MOPHOC, a mobile application to support the learning process of phonetic code. This tool allows to conduct several kinds of exercises related with articulation (phonemic awareness, sentence construction, phonatory, …).

User interface

This layer contains the bindings to connect the ICT tools (Intelligent ICT environment) with the representation and management knowledge layer. Other important elements of this layer are the speech-language therapy templates generated using OpenGenEHR tool (http://www.openehr.org). These templates provide the interface to feed the different information structures related with clinical data. An example of a template to access personal data, conduct a hearing screening, conduct the PTA test, and other options is presented in Figure 3 (b).

Results

Once we have deployed the knowledge model, the information exchange mechanisms, and the main algorithms of the expert system, three pilot experiments were carried with the aim to validate the developed ICT tools with the support of five institutions of special education from Ecuador: Instituto de Parálisis Cerebral del Azuay (Institute of Cerebral Palsy of Azuay), Unidad Educativa Especial del Azuay (Unit for Special Education of Azuay), Fundación "General Dávalos" (General Davalos Foundation), Fundación "Jesús para los niños" ("Jesus for the kids" Foudation), and CEDEI School. These institutions are concerned about the health care, education, and rehabilitation of children suffering from different disabilities (cerebral palsy, dysarthria, dyslalia, athetosis, autism, Down syndrome, multi-disabilities, ...). In this line, for the first experiment, a team of 4 experts were provided with the mobile application MOPHOC to conduct several phonetic exercises and evaluations using articulatory phonetic tests. A total of 32 children were evaluated, and the system determined the error produced in the utterance production of phonemes, consonant clusters, diphones, and sentences (Table 1).

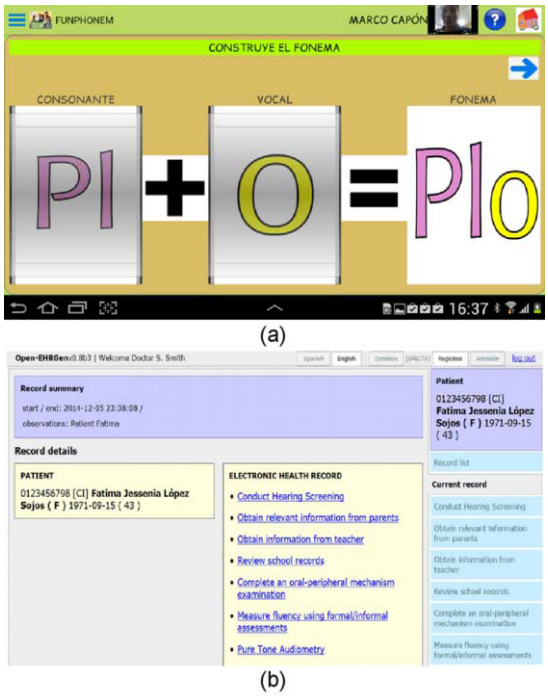


Figure 3– (a) A screen capture of the "articulation notebook" module of the MOPHOC application. (b) A screen capture from menu options of the archetype to conduct a speech evaluation for fluency

In this experiment, we also analyzed the time required to conduct the therapy and evaluation activities through the mobile application (MOPHOC). Also, we have verified an important reduction in the following activities: application of articulatory phonetics test (reduction from 30 to 15 minutes), planification of therapy activities using multimedia resources (5 minutes instead 30), and report generation (reduction from 30 to 5 minutes). These results have been validated by the experts, using manual version of PTA test.

Table 1– Most common articulatory errors detected by the system in the second group of children (32 cases)

Utterance	Average errors committed (%)
Consonant cluster	83
Sentence	77
Diphone	77
Phoneme	67

For the second experiment, a team of 6 experts was provided with the SPELMAT system to conduct evaluations on a different group of 53 children. This team of speech and language therapists had created 40 therapy plans to train the system, according to each child's profile. Each plan consists of five sub-plans to provide general therapy guidelines in 5 SLT areas: hearing, receptive language, linguistic formulation, oral structure and function, and expressive language and articulation, respectively (a total of 200 elementary plans). After that, the remaining 13 children's profiles were presented to the system, with the aim to generate a therapy plan for each (65 sub-plans). Therapists evaluated the outputs by giving each plan a rating of 1 if it was convenient and 0 otherwise, considering consistency, accurateness, and completeness. The last parameters were evaluated in the same way (1=correct, 0=incorrect); therefore, only if the three are correct, the accuracy will be 1. Table 2 shows the accuracy of SPELTA according to each area.

Table 2– Accuracy achieved in the generation of master therapy plans

Speech and language area	Plans correctly generated	Accuracy (%)
Hearing	13	100
Linguistic formulation	13	100
Oral structure and function	13	100
Receptive language	11	84
Expressive language+articulation	9	70
	Average	90

In the last experiment, we have conducted several activities to evaluate the robotic assistant as support tool for SLT. Some of the main goals were to analyze the response time of children during the relaxation activities and initial therapy, to test the robot response to the stimulation of motor skills, and to verify the robot's integration with the database of profiles and therapy plans. Some of the most relevant achievements were that the time needed to conduct the relaxation activities and initial therapy has shrunk from 40 minutes to 25, all the specialists have agreed on the usefulness of the robot, and children have shown high levels of motivation during their interactions with the robot.

Discussion

The results obtained with this study show that it is possible to automate several activities related with SLT, with the aim to provide a better service to patients suffering from several kinds of disabilities. In the same way, the use of mobile support tools allows to reduce significatively the time needed to conduct several activities related with therapy planification, assessment, and report generation. With these mobile tools, an SLP can perform several tasks from any place and assist his/her patients in a better way. Likewise, the use of expert systems based on clustering, allows SLP to discover "hidden populations" for those cases belonging to a group of subjects that suffer from the same disorders and have the same development cognitive age, but have deteriorations in different skills of the same speech-language area.

The management of electronic health record in conjunction with the concepts modeling based on ontologies have an important potential to represent the complex underlying relationships in SLT. With the use of these artifacts, it is possible to create knowledge databases for specific domains and provide efficient integration services for ICT tools and intelligent systems. The use of a model based on several layers, among other elements (archetypes, templates, ICT tools, etc.), provides a robust growing environment, able to easily incorporate new knowledge mainstays, actors, and supporting tools for sharing information through standardized languages and vocabularies.

Conclusion

In this paper we have presented a comprehensive model to support the most relevant task related with SLT. This model is designed to be adjusted or grow up according to requirements of the institutions where patients receive therapy and

rehabilitation. In the same way, given that our model uses archetypes, templates and standardized vocabularies, all generated information can be shared and updated in an efficient way.

Likewise, using this model, it is possible to improve several process related to management of SLT information, given that a SLP can generate reports, visualize data, and analyze the patients' progress quickly and from any device (web, mobile or desktop). On the other hand, the expert system layer provides support in decision making for complex tasks, such as planning and scheduling of therapy sessions and activities according to specificity of each patient's profile. The SPELTA has achieved promising results, given that the generated therapy plans were deemed consistent, accurate, and complete by the experts. Only the areas of receptive and expressive language seem to demand a refined or alternative approach.

The following are proposed lines of future work:

- Extend the proposed model to cover some deeper areas of knowledge, like the automatic generation of specific therapy plans based on daily activities and considering these elements: the existing levels of granularity of the patient's cognitive development, the incidence of other disorders (cerebral palsy, athetosis ...), etc.
- Modify the distance measures to improve the accuracy of generated plans in the areas of receptive language, and expressive language and articulation.

Acknowledgements

The authors from the Universidad Politecnica Salesiana have been supported by the “Sistemas Inteligentes de Soporte a la Educación CIDII-010213” research project. The authors from the University of Vigo have been supported by the European Regional Development Fund (ERDF) and Xunta de Galicia under project CN 2012/260 “Consolidation of Research Units: AtlantTIC”, and by the Ministerio de Educación y Ciencia (Gobierno de España) research project TIN2013-42774-R (partly financed with FEDER funds). We would like to thank the support provided by the following institutions of special education: Instituto de Parálisis Cerebral del Azuay (IPCA), Unidad Educativa Especial del Azuay (UNEDA), CEDEI School and Fundación “General Dávalos”.

References

[1] Conti-Ramsden G, and Durkin K. Language development and assessment in the preschool period. In: Neuropsychology review. Springer-Verlag, 2012.

[2] Poter R. *The Merck Manual of Diagnosis and Therapy*. 19th ed.: Merck, 2011.

[3] Hillis A (Ed.). *The handbook of adult language disorders*. Psychology Press, 2013.

[4] McAllister L, Wylie K, Davidson B, and Marshall J. The World Report on Disability: an impetus to reconceptualize services for people with communication disability. International journal of speech-language pathology 2013: 15(1): 118-26.

[5] Marshall J, Roper A, Galliers J, Wilson S, Cocks N, Muscroft S, and Pring T. Computer delivery of gesture therapy for people with severe aphasia. Aphasiology 2013: 27(9): 1128-46.

[6] Higgins C, Kearns A, and Franklin S. The development of a semantic feature analysis based mobile application for individuals with aphasia. Proceedings of the 10th international conference on Mobile systems, applications, and services, 2012;, pp. 513-4.

[7] Caballero-Morales SO, and Trujillo-Romero F. Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition. Expert Systems with Applications 2014: 41(3): 841-52.

[8] Loaiza D, Oviedo C, Castillo A, Portilla A, Alvarez G, Linares D, and Navarro A. A Video Game Prototype for Speech Rehabilitation. International Conference In Games and Virtual Worlds for Serious Applications (VS-GAMES), 2014; pp. 1-4.

[9] Alamoudi A, Almozaini M, Alabdulrahman R, Alkoblan S, Drine S, and Al-Wabil A. Interactive Serious Gaming for Children with Auditory Processing Difficulties in the Arabic Language. HCI International 2013-Posters’ Extended Abstracts, Springer Berlin Heidelberg, 2013.

[10] Martín Ruiz ML, Valero Duboy MÁ, Torcal Loriente C, and Pau de la Cruz I. Evaluating a Web-Based Clinical Decision Support System for Language Disorders Screening in a Nursery School. Journal of medical Internet research, 2014.

[11] Robles-Bykbaev V, López-Nores M, Pazos-Arias JJ, and Arévalo-Lucero D. Maturation Assessment System for Speech and Language Therapy based on Multilevel PAM and KNN. Procedia Technology, 2014; pp. 1265-70.

[12] American Speech-Language-Hearing Association. International Classification of Diseases, Tenth Revision - Clinical Modification, Related to Speech, Language, and Swallowing Disorders, 2014.

[13] Carter M, and Stephenson J. The use of multi-sensory environments in schools servicing children with severe disabilities. Journal of Developmental and Physical Disabilities 2012:24(1): 95-109.

[14] Poza J, Gómez C, Gutiérrez MT, Mendoza N, and Hornero R. Effects of a multi-sensory environment on brain-injured patients: Assessment of spectral patterns. Medical engineering & physics, 2013; pp. 365-75.

Address for correspondence

Vladimir Robles-Bykbaev
Universidad Politécnica Salesiana, Cuenca – Ecuador.
Calle Vieja 12-30 y Elia Liut
vrobles@ups.edu.ec

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-55

Identification of Patient Safety Risks Associated with Electronic Health Records: A Software Quality Perspective

Luiz A. Virginio Jr.[a], Ivan Luiz Marques Ricarte[b]

[a] *School of Electrical Engineering and Computation, University of Campinas, Campinas, SP, Brazil*
[b] *School of Technology, University of Campinas, Limeira, SP, Brazil*

Abstract

Although Electronic Health Records (EHR) can offer benefits to the health care process, there is a growing body of evidence that these systems can also incur risks to patient safety when developed or used improperly. This work is a literature review to identify these risks from a software quality perspective. Therefore, the risks were classified based on the ISO/IEC 25010 software quality model. The risks identified were related mainly to the characteristics of "functional suitability" (i.e., software bugs) and "usability" (i.e., interface prone to user error). This work elucidates the fact that EHR quality problems can adversely affect patient safety, resulting in errors such as incorrect patient identification, incorrect calculation of medication dosages, and lack of access to patient data. Therefore, the risks presented here provide the basis for developers and EHR regulating bodies to pay attention to the quality aspects of these systems that can result in patient harm.

Keywords:

Electronic Health Records; Patient Safety; Quality Improvement.

Introduction

The progress of information technology has impacted the healthcare sector. Some of these impacts are caused by adoption of the Electronic Health Records (EHR) [1]. One of the main goals of EHR is to support continuity, efficiency, and quality in healthcare [2]. These systems can offer benefits, such as ease of access to patient data, research support [3], and greater completeness and documentation comprehensiveness [2]. In addition, these systems have shown a capacity for reducing medical errors and increasing patient safety, mainly by means of decision-making support mechanisms [1,4].

Despite these benefits, the literature also presents evidence that, when developed and/or used improperly, EHR can incur risks to patient safety [5–8]. Patient safety is understood as a reduction of damage risks to patients in the healthcare process [9].

For example, some real cases of problems associated with the use of EHR that jeopardized patient safety can be cited. In 2007, a maintenance error in the network configuration of the Veterans' Health Administration (VHA), one of the largest healthcare providers in the United States, caused the EHR to become inaccessible for more than nine hours. As a result, many consultations were conducted without access to any documentation, and surgeries had to be postponed because the doctors were uncertain about how to proceed without proper documentation [7].

An evaluation of adverse event database from the Food and Drug Administration (FDA) in the United States also revealed cases of incidents related to the use of EHR. Loss or corruption of data, association of information to the wrong patient, and lack of access to the EHR resulted in problems such as delays in diagnosis or treatment, incorrect administration of medication, and even death [5].

Despite the growing quantity of such evidence, there is a significant gap of initiatives that address patient safety issues associated with the use of EHR [10]. Therefore, there is a need for more research aimed at investigating the possible EHR quality problems that can negatively affect the healthcare process.

In this context, this work presents a literature review of the risks that EHR can cause regarding patient safety. The risks presented here are classified based on the software product quality characteristics specified by the 25010 standard by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) [11]. This classification allows EHR developers to have better traceability of the quality characteristics of their products that can affect patient safety. According to the authors' knowledge, such an approach has not been used by any published work in the literature.

Methods

In order to search the relevant databases, we used medical subject headings (MeSH) terms and free terms defined from the two main topics associated with the theme of this work: "EHR" and "Patient Safety Risks." The electronic databases consulted were PubMed, IEEE Xplore Digital Library, ACM Digital Library, and ScienceDirect. We selected only papers published between 2010 and 2014.

The electronic database searches returned 8,609 references, 17 of which were selected for review. The works were selected by an iterative scan of these references in order to eliminate duplicate records and select only those references relevant to the purpose of this work.

After the literature review, the selected articles were submitted to an in-depth reading for risk identification. The risks were then classified based on the software product quality model established by ISO/IEC 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. This standard establishes a model composed of eight characteristics (divided into sub-characteristics) used to evaluate software product quality and evaluate how it can meet the needs of stakeholders [11].

Results

The works included in this literature review investigated the risks to patient safety associated with EHR using approaches such as: analysis of incident reports [5,12–16], application of interviews or questionnaires to experts or EHR users [17–19], literature review [20–23], and expert opinion [24–26].

In the next sections, we present the risks to patient safety associated with EHR identified by reading the articles included in the literature review. These risks are presented based on the characteristics and sub-characteristics of software product quality set out in ISO/IEC 25010, as listed in Table 1.

Table 1– Risks classified by ISO/IEC 25010 software quality model

Characteristics	Sub-characteristics	Risk
Functional Suitability	Functional completeness	R1 - Lack of functionalities that support clinical workflow [5,17,20]
		R2 - Lack of coding, standardization, and structuring data [24]
		R3 - Lack of features to detect duplicate patient records [19,27]
	Functional correctness	R4 - Inaccurate, outdated, or incomplete decision support rules [5,20,21]
		R5 - Software bugs [5,13,14,20,23]
	Functional appropriateness	R6 - Inadequate content import features [20,21,23,26]
		R7 - Pre-populated fields [20,23,25,26]
		R8 - Inadequate alerting [19,20,22,24]
		R9 - Allow tasks to be performed simultaneously [12,19]
Usability	Appropriateness recognizability	R10 - Inadequate display of information [5,12,14,16,18,19,22,25]
	Learnability	R11 - Difficulty in understanding current status of user actions [19]
	Operability	R12 - Difficulty in interacting with EHR [19,21]
	User error protection	R13 - Interface prone to user error [12,16,17,19–22,27]
	User interface aesthetics	-
	Accessibility	-
Performance Efficiency	Time behavior	R14 - Delay in system response [12]
	Resource utilization	-
	Capacity	-
Compatibility	Co-existence	-
	Interoperability	R15 - Communication errors between systems [5,12,14–16,19,24]
Reliability	Availability	R16 - EHR unavailability [13,15,17–19,24,25]
	Maturity	-
	Fault tolerance	-
	Recoverability	-
Security	Confidentiality	-
	Integrity	-
	Non-repudiation	-
	Authenticity	-
	Accountability	-
Maintainability	Modularity	-
	Reusability	-
	Analyzability	-
	Modifiability	-
	Testability	-
Portability	Adaptability	-
	Installability	-
	Replaceability	-

Functional Suitability

The quality characteristic "functional suitability" is related to the degree to which the features offered by the software are complete, accurate, and appropriate. Risks associated with all three sub-characteristics in this category were identified.

Functional Completeness

The sub-characteristic "functional completeness" is the degree to which the software provides necessary functions for users to achieve their goals. The following risks have been identified for this category:

- **R1** - Lack of functionalities that support clinical workflow [5,17,20].
- **R2** - Lack of coding, standardization, and structuring data [24].
- **R3** - Lack of features to detect duplicate patient records [19,27].

The absence of features that support clinical workflow can cause users to decide on workarounds that could jeopardize patient safety. For example, when the EHR does not allow registration of drug administration prior to registering its order (a common necessity in emergencies), the drug application documentation will occur after administration, which can result in medication being administered again because its administration was not registered previously [20].

The lack of codification, standardization, and data structure can result, for example, in failure to issue alerts [24], whereas the absence of mechanisms to detect double-patient records can result in documentation gaps because of information fragmentation [27].

Functional Correctness

The sub-characteristic "functional correctness" evaluates the degree of accuracy and correctness of the functionalities offered by the software. For this category, the following risks were found:

- **R4** - Inaccurate, outdated, or incomplete decision support rules [5,20,21].
- **R5** - Software bugs [5,13,14,20,23].

Supporting rules for incorrect, not current, or incomplete decisions can result in the issuance of a large amount of false-positive alerts, encouraging users to overlook alerts [21]. In addition, users with high confidence in technology can be mislead when making decisions based on incorrect recommendations issued by the system. This process is known as "automation bias," in which users perform actions recommended by technology when they have decision doubts, or even when such actions contradict their knowledge [20].

Software bugs can cause incorrect dosage calculations [5] or even the corruption, loss, or improper storage of data [20]. This risk was also associated with maintenance or updates to EHR that, when poorly managed, can introduce new bugs to the software [14].

Functional Appropriateness

The sub-characteristic "functional appropriateness" verifies how the functions offered by the software are appropriate and capable of facilitating task execution by users. The following risks were identified:

- **R6** - Inadequate content import features [20,21,23,26].
- **R7** - Pre-populated fields [20,23,25,26].
- **R8** - Inadequate alerting [19,20,22,24].
- **R9** - Allow tasks to be performed simultaneously [12,19].

Text import features (such as copy/paste) are often related to the propagation of incorrect information, loss of authority assignment, or even copy of outdated information [26]. Whereas these problems occur mainly from abusing the use of these features [17,20], the EHR can help avoid them by restricting the type of data that can be imported, and including the original text source for imported content [26].

Fields with default values (such as the establishment of standard doses [23,25]) can also incur risks to patients when not reviewed by users [26]. On the other hand, issuing alerts and reminders with low specificity or sensitivity, irrelevant, or excessive, encourages users to overlook alerts that could potentially be important [19,22] and interrupts the clinical workflow and thought process of health professionals [24],.

Allowing certain tasks to be performed simultaneously, such as opening two or more patient records on the same device [19] or simultaneously editing the same record by different users [12], can result in the registration of information to the wrong patient, or in information inconsistency [19].

Usability

The quality characteristic "usability" evaluates user interface in the context of ease of understanding, learning, ease of use, user attraction, and accessibility. For this category, risks were found in the sub-characteristics "appropriateness recognizability," "learnability," "operability," and "user error protection".

Appropriateness Recognizability

The sub-characteristic "appropriateness recognizability" verifies the degree to which users recognize the software suitable for their needs. The following risk for this category was identified:

- **R10** - Inadequate display of information [5,12,14,16,18,19,22,25].

The problems related to this risk are in regards to incomplete display of information [5,12,18,19,22,25], such as not displaying pre-existing medications or patient allergy data [12]; buttons with the same label, but different features [25]; and presentation of high information volume [5]. These problems are associated with patient misidentification (caused by not displaying key identifying data) [19] and incorrect interpretation of information [16].

Learnability

The sub-characteristic "learnability" evaluates how the software allows users to understand key software concepts, thus making the concepts effective for use. The following risk was identified:

- **R11** - Difficulty in understanding current status of user actions [19].

This risk is associated with the occurrence of open or incomplete orders caused by the failure to complete an order process, including signature or submission. This problem can be caused by the user interface when it becomes difficult to understand the current status of user's actions [19].

Operability

The sub-characteristic "operability" assesses the ease of using the software. The following risk was found for this category:

- Modelo de Qualidade ISO/IEC 25010 - Difficulty in interacting with EHR [19,21].

Interface problems, such as the display of information of the same context on multiple screens or tabs [19,21], can cause user interaction with EHR difficult. Difficulty in navigating, visualizing, understanding, and interacting with the user interface can cause failure to identify and/or use the most recent patient data, thus causing clinical decision errors and generating incorrect, unnecessary, or delayed tests, procedures, and therapies [19].

User Error Protection

The sub-characteristic "user error protection" evaluates the software's ability to prevent user errors. The following risk was identified:

- **R13** - Interface prone to user error [12,16,17,19–22,27].

Many of the selected studies reported that the way in which the user interface arranges information can facilitate the occurrence of user errors. The main problems regarding this risk are related to item selection lists (such as drug lists) or drop-down menus that often have very similar items [19–21], lack of item grouping [19], or a very large amount of items [19,21]. These design issues can result in the improper

selection of items (such as incorrect patient selection) [16,17,19,27].

Performance Efficiency, Compatibility, and Reliability

The quality characteristic "performance efficiency" evaluates software optimization in connection with processing time, resource consumption, and processing capacity. For this category, only the sub-characteristic "time behavior" presented a risk:

- **R14** - Delay in system response [12].

A delay in response to a particular EHR action can cause users to click the same action several times, thus generating duplicate prescriptions, for example [12].

The quality characteristic "compatibility" refers to the degree to which two or more systems or components of a system can exchange information. One risk was identified for the sub-characteristic "interoperability":

- **R15** - Communication errors between systems [5,12,14–16,19,24].

Communication errors between systems or components of a system are often caused by difficulties in EHR interoperability with other systems [22], or failures in network infrastructure [5,12,14,16,25]. Such errors can prevent patient context and status from being consistent between components/systems, leading mainly to delays in the healthcare process [14].

The quality characteristic "reliability" evaluates software in terms of maturity (frequency with which the software is defective over time), availability (degree to which the software is available for use), fault tolerance (the software's ability to operate in the presence of hardware or software failures), and recoverability (the software's ability to recover data and return to operate after the occurrence of a failure). For this category, we found one risk with respect to the sub-characteristic "availability":

- **R16** - EHR unavailability [13,15,17–19,24,25].

When EHR is not available for use, health professionals can lose access to the documentation required for healthcare processes, which can lead to such problems as delays in diagnosis and treatment [7].

The main causes of lack of access to EHR are failures or planned downtime of some infrastructural component [19]. Therefore, it is necessary for EHR to offer redundant devices [19,24], and a read-only version to access without connection [19].

Discussion

The software quality characteristics that presented more risks to patient safety were "functional suitability" (nine risks) and "usability" (four risks). On the other hand, the risks most discussed in the articles selected for this literature review were: "Inadequate display of information" (usability-appropriateness recognizability); "interface prone to user error" (usability-user error protection); "communication errors between systems" (compatibility-interoperability); and EHR unavailability (reliability-availability).

Once the major quantity of identified risks is related to the functions provided by the EHR (functional suitability) and its usability, it is noticed the importance of the integration of the end user to the product development process. A way for achieving such integration is the use of the principles of User-Centered Design (UCD) [28]. The UCD is an approach that involves the end users along all of the development process in order to enssure that the product is suitable to their needs. This approach also seeks to take into account the needs and expectations of any person that could be affected by the use of the product [29] (e.g. the patients). Therefore, UCD may help to building EHRs more compliant to the health professionals' needs and, at the same time, safer for patients.

No risks were found for the quality characteristics "security," "maintainability," and "portability." However, the authors consider that the attribute "security" (which evaluates how the software can protect its functions and data from unauthorized access) can also be important for patient safety because problems regarding confidentiality breaches of patient information could harm their psychological well-being.

In addition to the risks specifically related to EHR quality characteristics, we identified risks that arise because of improper EHR use, such as neglecting the use of structured fields to use only free text fields [24]; infrastructure problems, such as network failures [14]; and other contributing factors, such as gaps in regulations [20] and poor user training [16,23].

Therefore, it is clear that the EHR is part of a complex sociotechnical system in which patient safety incidents can occur because of the interaction of several factors [30]. However, this study aimed to present risks in an approach that was more focused on EHR quality characteristics, and therefore, risks relating to other aspects are not discussed.

Conclusion

This paper presented risks that EHR can incur to patient safety when developed improperly. The risks were presented according to the characteristics of software product quality set out in ISO/IEC 25010, thus allowing the developers of these systems to have a reference for identifying the quality attributes of their products that might pose risks to patient safety.

In addition, the risks presented can be used as reference for the specification of more stringent requirements in the EHR certification processes, once these processes still require more direct approaches to patient safety [10,31].

However, it is still necessary to investigate strategies to mitigate the risks presented here in order to provide techniques and recommendations that can be used by EHR developers to create a product that offers greater safety to the healthcare process.

It is also necessary to examine the risks to patient safety in a sociotechnical context in greater depth, thus allowing the identification of problems related to EHR process use, specification of regulations, supervision, infrastructure, and any other factors that interact with EHR throughout its lifecycle. The identification of these problems should be followed by the identification of their solutions, indicating the role of each actor involved in this sociotechnical system in mitigating the risks to patient safety.

References

[1] Chaudhry B. Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Ann Intern Med*. 2006;144(10):742.

[2] Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*. 2008;77(5):291–304.

[3] Hoffman S, Podgurski A. Finding a cure: the case for regulation and oversight of electronic health record

systems. *Case Leg Stud Res Pap No 08-13; Harvard J Law Technol*. 2008;22(1).

[4] Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, Taylor R. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff*. 2005;24(5):1103–17.

[5] Myers RB, Jones SL, Sittig DF. Review of Reported Clinical Information System Adverse Events in US Food and Drug Administration Databases. *Appl Clin Inform*. 2011;2(1):63–74.

[6] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors. *J Am Med Assoc*. 2005;293(10):1197–203.

[7] Johnson CW. Politics and patient safety don't mix: understanding the failure of large-scale software procurement in healthcare. *IET Conf Proc*. Institution of Engineering and Technology; 2009;33–33(1).

[8] Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. *J Am Med Inform Assoc*. 2007;14(5):542–9.

[9] Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an International Classification for Patient Safety: key concepts and terms. *Int J Qual Health Care*. 2009;21(1):18–26.

[10] Magrabi F, Aarts J, Nohr C, Baker M, Harrison S, Pelayo S, Talmon J, Sittig DF, Coiera E. A comparative review of patient safety initiatives for national health information technology. *Int J Med Inform*. Elsevier Ireland Ltd; 2013;82(5):e139–48.

[11] Internation Organization for Standardization, International Electrotechnical Commission. ISO/IEC 25010:2011 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models [Internet]. Internation Organization for Standardization; 2011. p. 34.

[12] Magrabi F, Ong M-S, Runciman W, Coiera E. An analysis of computer-related patient safety incidents to inform the development of a classification. *J Am Med Informatics Assoc*. 2010;17(6):663–70.

[13] Lei J, Guan P, Gao K, Lu X, Chen Y, Li Y, Meng Q, Zhang J, Sittig DF, Zheng K. Characteristics of health IT outage and suggested risk management strategies: an analysis of historical incident reports in China. *Int J Med Inform*. Elsevier Ireland Ltd; 2014;83(2):122–30.

[14] Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. *J Am Med Informatics Assoc*. 2014;1–7.

[15] Warm D, Edwards P. Classifying health information technology patient safety related incidents - an approach used in Wales. *Appl Clin Inform*. 2012;3(2):248–57.

[16] Magrabi F, Ong M-S, Runciman W, Coiera E. Using FDA reports to inform a classification for health information technology safety problems. *J Am Med Informatics Assoc*. 2012;19(1):45–53.

[17] Meeks DW, Takian A, Sittig DF, Singh H, Barber N. Exploring the sociotechnical intersection of patient safety and electronic health record implementation. *J Am Med Informatics Assoc*. 2014;21(e1):e28–34.

[18] Menon S, Singh H, Meyer AND, Belmont E, Sittig DF. Electronic health record-related safety concerns: A cross-sectional survey. *J Healthc Risk Manag*. 2014;34(1):14–26.

[19] Sittig DF, Singh H. A red-flag-based approach to risk management of EHR-related safety concerns. *J Healthc Risk Manag*. 2013;33(2):21–6.

[20] Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Heal Inf Manag*. 2013;10:1c.

[21] Khajouei R, Jaspers MWM. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: a systematic review. *Methods Inf Med*. 2010;49(1):3–19.

[22] Lawler EK, Hedge A, Pavlovic-Veselinovic S. Cognitive ergonomics, socio-technical systems, and the impact of healthcare information technologies. *Int J Ind Ergon*. Elsevier Ltd; 2011;41(4):336–44.

[23] Mangalmurti SS, Murtagh L, Mello MM. Medical malpractice liability in the age of electronic health records. *N Engl J Med*. 2010;363(21):2060–7.

[24] Sittig DF, Singh H. Electronic health records and national patient-safety goals. *N Engl J Med*. 2012;367(19):1854–60.

[25] Sittig D, Singh H. Defining health information technology–related errors: New developments since To Err Is Human. *Arch Intern Med*. 2011;171(14):1281–4.

[26] Weis JM, Levy PC. Copy, paste, and cloned notes in electronic health records: prevalence, benefits, risks, and best practice recommendations. *Chest*. 2014;145(3):632–8.

[27] McCoy AB, Wright A, Kahn MG, Shapiro JS, Bernstam EV, Sittig DF. Matching identifiers in electronic health records: implications for duplicate records and patient safety. *BMJ Qual Saf*. 2013;22(3):219–24.

[28] Internation Organization for Standardization. ISO 9241-210:2010 - Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems [Internet]. 2010.

[29] Abras C, Maloney-Krichmar D, Preece J. User-centered design. *Bainbridge, W Encycl Human-Computer Interact Thousand Oaks Sage Publ*. 2004;37(4):445–56.

[30] Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Qual Saf Health Care*. 2010;19 Suppl 3:i68–74.

[31] Hoffman S, Podgurski A. Meaningful use and certification of health information technology: what about safety? *J Law Med Ethics*. 2011;39 Suppl 1:77–80.

Address for correspondence

Luiz A. Virginio Jr.

E-mail: luiz.virginio.jr@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-60

Nonadherence to Oral Antihyperglycemic Agents: Subsequent Hospitalization and Mortality among Patients with Type 2 Diabetes in Clinical Practice

Vivienne J. Zhu[a], Wanzhu Tu[b], Marc B. Rosenman[a,b], J. Marc Overhage[c]

[a]Regenstrief Institute, IN, USA [b]Indiana University School of Medicine, IN, USA, and [c]Cerner Corporation, Malvern, PA, USA

Abstract

Using real-world clinical data from the Indiana Network for Patient Care, we analyzed the associations between non-adherence to oral antihyperglycemic agents (OHA) and subsequent diabetes-related hospitalization and all-cause mortality for patients with type 2 diabetes. OHA adherence was measured by the annual proportion of days covered (PDC) for 2008 and 2009. Among 24,067 eligible patients, 35,507 annual PDCs were formed. Over 90% (n=21,798) of the patients had a PDC less than 80%. In generalized linear mixed model analyses, OHA non-adherence is significantly associated with diabetes related hospitalizations (OR: 1.2; 95% CI [1.1,1.3]; p<0.0001). Older patients, white patients, or patients who had ischemic heart disease, stroke, or renal disease had higher odds of hospitalization. Similarly, OHA non-adherence increased subsequent mortality (OR: 1.3; 95% CI [1.02, 1.61]; p<0.0001). Patient age, male gender, income and presence of ischemic heart diseases, stroke, and renal disease were also significantly associated with subsequent all-cause death.

Keywords

Health information exchange, type 2 diabetes, medication adherence, hospitalization, mortality

Background

Type 2 diabetes is a major public health crisis in this country. Over 29 million adults in the United States currently have diabetes mellitus, and the incidence is increasing at nearly epidemic rates [1]. This will have a significant public health impact because diabetes is a major cause of morbidity and mortality. In addition, healthcare costs are 11% higher for patients with poorly controlled diabetes compared to those with good glycemic control [2].

Clearly, pharmacotherapy for glycemic control is essential to achieving the goal of long-term metabolic control and reducing the risk of cardiovascular disease (CVD) events, the primary cause of hospitalization and mortality among these patients. However, increasing evidence suggests that patients with diabetes often have poor adherence to prescribed medication therapies. A systematic review of medication adherence in diabetes showed low adherence rates, ranging between 36-87% across numerous studies [3]. These observations have been replicated in several studies, and adherence among patients with diabetes is disappointingly low, dropping most dramatically after the first six months of therapy [4,5].

Improving medication adherence is viewed as a high priority for health care reform. The Centers for Medicare and Medicaid Service (CMS) has encouraged adherence to oral diabetes medications via quality and efficiency measures in health and drug plan performance rating [6]. However, improving medication adherence is a multifactorial challenge. Health information technology (HIT) and health information exchange (HIE) are playing an increasing role in generating, measuring, and transferring important medication adherence data between hospitals, pharmacies, providers, and patients. Recently, the electronic medical record system (EMR) meaningful use criteria and the Accountable Care Organization (ACO) have encouraged electronic prescribing, medication reconciliation, patients' accessibility to their own medical information, and care coordination [7,8], all of which support improving medication adherence and ultimately improving patient health outcomes through HIT and HIE.

Although the effects of medication adherence on patient hospitalization and mortality, as well as the significant potential of HIT and HIE for improving medication adherence, have been increasingly recognized, few population-based studies have been conducted using real-world data. The association between medication non-adherence and mortality has been chiefly demonstrated in clinical trials. In this study, we aimed to assess adherence to OHA for patients with type 2 diabetes and to analyze the association between OHA non-adherence and hospitalization and mortality, using real-world clinical data from an operational HIE.

Methods

Data sources and settings

This retrospective study was designed to analyze the association between nonadherence to oral antihyerglycemic agents (OHA) and health outcomes in the subsequent year for patients with type 2 diabetes. We used medication dispensing data from 2008 to 2009 and clinical data (hospitalization and death) from 2009 to 2010 from the Indiana Network for Patient Care (INPC). The INPC is an operational, regional, health information exchange that included over 100 hospitals, physician practices, payers, pharmacies, laboratories, and independent radiology centers. This system delivers medical record information from hospitals, laboratories, imaging centers, pharmacies and physician offices [9]. In particular the INPC included pharmacy claims data from the largest public and private payers as well as medication history data from pharmacy benefit managers obtained via the Surescripts network [10]. For this study, the patient medication history, diabetes related hospitalizations, and death (all causes) information were extracted from the INPC and were linked through a robust linkage algorithm. We additionally estimated patient income through median household income by zip code obtained from the U.S. Census Bureau website (2000 data is the most recently information available). This

study was approved by the Institutional Review Board of Indiana University and the INPC Management Committee.

Eligibility criteria

Eligible subjects were age 18 years or older in 2008, and had a diagnosis of type 2 diabetes using the *International Classification of Disease, Ninth Revision, Clinical Modification* (ICD-9_CM) of 250.x0 and 250.x2. They had at least one pharmacy claim for an OHA: Biguanides, Sulfonylurea (SU), Thiazolidinedione (TZD), other OHAs (Meglitinides and α- glucosidase), or OHA combination. Patient OHA dispensing information was extracted by the National Drug Codes (NDC) with drug names matched to OHA medications identified by the First DataBank *Standard Therapeutic Code* system (STC:71, excluding injectable: 0177, A771, A716), from 2008 to 2009. Patients who used insulin were excluded. This sample reflects patients who had a drug benefit. Also, since all included subjects had in fact filled a prescription, it suggests that they used that benefit. Therefore, it is likely that the data collected accurately reflects the pattern of medication dispensing events.

Measurements

The independent variable is the annual OHA medication adherence for 2008 and 2009, which was measured using a standard approach: the proportion of days covered (PDC). The PDC is an accepted standard measurement for evaluating medication adherence using retrospective data. It is defined as the total number of medication-covered days divided by the number of days in a certain time period. In this study, the annual PDC is calculated as the proportion of days which had at least one OHA available during a 365-day period. Medication history data, including refill date, days of supply, dosage, and frequency, are used to calculate PDC [11]. For patients who took multiple OHAs, we calculated the combined PDC. Non-adherence was defined as a PDC less than 80%, a conventional cut- off point to define poor adherence for chronic conditions [12].

The dependent variables were diabetes hospitalization and all-cause death for 2009 and 2010

(2010 is the most recent year for which death certificate data are available in the INPC). For each patient, we identify whether a diabetes-related hospitalization occurred in 2009 or 2010 according to the primary diagnosis for inpatient admissions in the INPC. Patients who had at least one recorded hospitalization related to diabetes type 2 (ICD-9: 250.*2), cardiovascular disease (ICD-9:410.*. 411.*, 412.*, 414.*), stroke (ICD-9:433.*1, 434.*), or renal disease (ICD-9:585.*) during a study year were considered to have a DM2 related hospitalization. Similarly, the patient's vital status also was identified from the INPC which was previously integrated from the Social Security Administration.

In order to control possible confounders that may influence patient hospitalization and mortality, we examined age, gender, race, income, number of concurrent OHAs (single, multiple), and baseline co-morbidities (hypertension, ischemic heart diseases, stroke, and renal disease) as covariates.

Statistical analyses

Patient OHA adherence was assessed. Descriptive statistics of patient demographic and clinical characteristics, hospitalizations, and mortality were reported for both the adherent and non-adherent groups. Continuous variables were compared using *t*-tests, and categorical variables were compared using $\chi 2$ tests. Generalized linear mixed models (GLMMIX) were performed to evaluate the association between hospitalization or mortality and OHA adherence of the previous year in the INPC DM2 population. Random subject effects were used in these models to accommodate the potential association among observations contributed by the same study subjects. Adjusted odds ratios (OR) were used to quantify the associations. All analyzes were implemented using SAS 9.1 (*SAS Institute, Cary, North Carolina*). Two-sided *p*-values less than 0.05 were considered significant.

Results

A total of 24,067 eligible patients with DM2 were identified for this study from 2008 to 2010, and 35,069 annual PDC were formed. Across the study period, 21,798 (90.6%) of the patients had a mean annual PDC of less than 80%. A total of 4,235 (17.9%) patients had at least one DM2 related hospitalization, and 319 (1.3%) patients died by 2010. Patient characteristics by adherence are outlined in Table 1. Non-adherent patients were relatively younger (60 vs 67), more were female (52% vs. 49%), fewer were white (81% vs 87%), and they had relatively lower income ($45,447 vs $46,872). Around 70% of patients had hypertension in both groups. The rates of hypertension and ischemic heart disease was slightly lower in the non-adherent group, but no significant difference was found for stroke and renal disease between the two groups.

Table 1– Patient characteristics

	Adherent (N=2,269)	Nonadherent (N=21,798)
	N (%)	N (%)
Age at study entry†	67.0 (12.1)	60.4 (15.1)
Female gender	1,114 (49.1)	11,369 (52.2)
Race		
Asian	7 (0.3)	52 (0.2)
African-American	289 (12.7)	3,869 (18.1)
Hispanic	4 (0.2)	158 (0.7)
White	1,969 (86.8)	17,628 (80.9)
Multiple OHA agents	683 (30.1)	6,771 (31.1)
Hospitalization	377 (16.6)	3,945 (18.1)
Death	25 (1.1)	294 (1.3)
Comorbidity		
Hypertension	1,665 (73.4)	15,163 (69.6)
Ischemic heart disease	570 (25.1)	4,893 (22.5)
Stroke	54 (2.4)	536 (2.5)
Renal disease	192 (8.5)	1,669 (7.7)
Median income ($) †	46,872 (9,303)	45,447 (12,654)

† mean (SD)

Table 2– Association between adherence, hospitalization, and death

	Hospitalization		Death	
	Odds Ratio (95%CI)		Odds Ratio (95%CI)	
Non-adherence	1.21	(1.12, 1.31)	1.28	(1.02, 1.61)
Age (10 yr incments)	1.06	(1.02, 1.20)	1.85	(1.10, 1.91
Female gender	1.14	(1.08, 1.31)	0.80	(0.65, 0.95)
Race (ref='White)				
Asian	0.19	(0.06, 0.60)	1.10	(0.16, 8.85)†
African-American	0.79	(0.73, 086)	0.75	(0.59, 0.96)
Hispanic	0.47	(0.29, 0.74)	0.71	(0.17, 2.90)†
Comorbidity				
Hypertension	1.01	(0.94, 1.07)†	0.65	(0.54, 0.79)
Ischemic heart disease	1.82	(1.70, 1.95)	1.24	(1.02, 1.52)
Stroke	1.56	(1.42, 1.71)	2.09	(1.68, 2.89)
Renal disease	1.42	(1.23, 1.64)	2.06	(1.49, 2.89)
Multiple OHAs	0.89	(0.82, 0.92)	0.96	(0.79, 1.15)†
Income ($10,000)	0.99	(0.92, 1.06)†	0.87	(0.82, 0.94)

† *p*-value >0.05

Compared with adherent patients, patients who did not adhere to OHA had significantly higher hospital admission rates (18.1% vs. 16.6%) and higher

mortality (1.35% vs. 1.1%). Table 2 summarizes the results of the GLIMMIX model that analyzed associations between OHA adherence, patient factors and subsequent DM2 related hospitalizations and mortality. After adjusting for patient baseline characteristics, OHA non-adherence is significantly associated with hospitalizations (OR: 1.2; 95% CI [1.1, 1.3]; p<0.0001). Older patients, whites, or patients who had ischemic heart disease, stroke, or renal disease had greater odds of hospitalization. Similarly, OHA non-adherence increased subsequent death due to any cause (OR: 1.3; 95% CI [1.0, 1.6]; p<0.0001). Patient age, male gender, and presence of a co-morbid condition (including ischemic heart diseases, stroke, or renal disease) were also significantly associated with subsequent death. Hypertension had no effect on hospitalization (p=0.84) but was associated with lower risk of death (OR; 0.5; 95% CI [0.54, 0.79]; p<0.001). Multiple OHA use decreased risk of hospitalizations, but had no significant association with death. Patient income had no association with hospitalization, but it was inversely associated with all-cause death.

Discussion

This large, population-based, observational study documented a high prevalence (90%) of non-adherence to OHA in clinical practice. In the multivariable GLIMMIX analyses, after fully adjusting for baseline patient demographic and clinical characteristics, non-adherent patients had a 1.2 fold higher risk for hospitalization with diabetes, cardiovascular diseases, or renal disease in the subsequent year. In addition, we observed significantly increased rates of all-cause mortality (OR: 1.3) among patients who did not adhere to OHA. These findings are consistent with previous population-based studies. Joe Hong, *et al.*, reported that non-adherence to OHA in the first two years increased the risk of hospitalization (OR: 1.26) and death (OR: 1.4) in the third year among patients in a national insurance program [13]; Lau and Nau found a 2.5 times higher risk for diabetes and cardiovascular disease hospitalization among non-adherent patients within one year using a registry [14]; Ho, et al., showed medication non-adherence was significantly associated with all-cause hospitalization (OR: 1.6) and all cause-mortality (OR: 1.8) in 12 months of follow up of an randomized clinical trial [15];

Glycemic control is the primary goal for diabetes management. Oral antihyperglycemic therapies are effective methods to control glycemic levels among patients with type 2 diabetes, thus lowering their risks of developing microvascular and macrovascular complications [16]; Previous studies have shown that OHA adherence was independently associated with HbA1c control: HbA1c decreases 0.10% to 0.16% for each 10% increment in OHA adherence [17]; As a result, improving OHA adherence may lead to better patient health outcomes through better glycemic control. Our study confirmed the importance of OHA adherence in clinical practice. Patients who did not adhere to their OHA medication (PDC less than 80%) were at higher risk of both hospitalization and mortality within one year.

In addition to the therapeutic effectiveness of OHA adherence, good adherence could also be a surrogate for other factors that reflect a high quality of care, healthy behavior, or more effective treatment. Patients who adhere to OHA may be more compliant with other medications, such as HMG-CoA reductase inhibitors and beta-blockers, to better control complications. They may also more actively implement self-management, including monitoring blood glucose levels, following a diet, engaging in regular exercise, caring for their feet, and interacting with providers, which all lead to improved clinical outcomes in diabetes. In contrast, non-adherence may be associated with depression, cognitive impairment, or missed appointments, which may be linked to less than desirable health outcomes.

This study also found that older patients and those with ischemic heart diseases, stroke, and/or renal diseases had higher risks of both hospitalization and death. Interventions for improving medication adherence or other intensive care should target these subgroups so that patients can achieve the full benefits of anti-hyperglycemic therapies. Interestingly, hypertension was not significantly associated with hospitalization and decreased risk of death, which is inconsistent with the notion that hypertension accelerates the progression of both micro- and macro-vascular complications in diabetes. A possible explanation could be that aggressive antihypertensive treatments, including a range of antihypertensive drugs and lifestyle therapies, might be applied to diabetic hypertensive patients to reduce the risk of micro-vascular and macro-vascular disease [18].

Studies about socioeconomic status and patient health outcomes have shown mixed results. Our study found no association between income and hospitalization, while patients with lower income had higher risk of all-cause death. The study subjects are mainly resident in the central Indiana region and have insurance coverage (including Medicaid), with which they might have been provided equal access to the healthcare system. However, low income may directly influence mortality by other context, such as health behaviors, stress, suicide, and homicide, as well as environmental factors [19].

The main findings from this study add evidence to the growing need to establish interventions in medication adherence in order to improve health outcomes for patients with type 2 diabetes. Barriers to medication adherence in diabetes exist at the patient, medication, and provider levels, and a multi-dimensional approach is required to establish efficient interventions. Health HIT and HIE offer the potential to establish such a system. First, objective and data-driven approaches can be programmatically established through an HIE. Second, computerized alerts and reminders can notify pharmacists or prescribers when their patients miss an opportunity to fill a medication in routine clinical practice. Third, predictive modeling may identify patients who are at higher risk of medication non-adherence and may help evaluate specific barriers for these patients. Fourth, widely used Internet and mobile health technologies may empower patients to manage their medications. In addition, a well-established HIE/HIT could support patient-centric and team-based care that better engages patients, providers, and health care systems for improving medication adherence.

Several limitations should be noted. First, the OHA dispensing information was extracted from the Surescript in the INPC covered geographic area; findings may not apply to other population. Second, dispensing claims might not reflect patient medication-taking behavior if patients did not actually use these medications. Nevertheless, filling a prescription is usually consistent with taking the medication. Third, findings from this observational study demonstrated associations (not causations) between OHA non-adherence and hospitalization, as well as associations (not causations) between OHA non-adherence and mortality. Fourth, some patients with type 2 diabetes may use insulin to control their glycemic levels; however, we lack information regarding insulin use for this population. In addition, we were not able to identify the specific cause of death using existing data; therefore, the association between OHA adherence and mortality might be overestimated.

Conclusions

This study found that 90% of patients with type 2 diabetes did not adhere to oral antihyperglycemic agents in clinical practice. Nonadherence was significantly associated with increased risks of hospitalization and death in the following year. The study findings emphasize the importance of improving medication adherence in diabetes management. Developing HIT and HIE strategies should be highly encouraged to effectively intervene with patients and providers regarding medication adherence.

References

[1] Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of diabetes and its burden in the United States, 2014. Atlanta, GA: U.S. Department of Health and Human Services; 2014.

[2] Hogan P, Dall T, Nikolov P. Economic Costs of diabetes in the US in 2002. Diabetes Care. 2003;26:917-32.

[3] Lee WC, Balu S, Cobden D, Joshi AV, Pashos CL. Prevalence and economic consequences of medication adherence in diabetes: a systematic literature review. Manag Care Interface. 2006 Jul;19(7):31-41.

[4] Mason BJ, Matsuyama JR, Jue SG. Assessment of sulfonylureas adherence and metabolic control. Diabetes Educ. 1995; 21(1):52-7.

[5] Rubin RR. Adherence to pharmacologic therapy in patients with type 2 diabetes mellitus. Am J Med. 2005 May;118 Suppl 5A:27S-34S.

[6] Centers for Medicare and Medicaid Services. Medicare health & drug plan quality and performance ratings 2012 Part C & Part D. 2011;45.

[7] Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010 Aug 5; 363(6):501-4.

[8] McClellan M, McKethan AN, Lewis JL, Roski J, Fishere ES. A national strategy to put accountable care into practice. Health Aff (Millwood). 2010 may; 29 (5): 982-90.

[9]. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, Mamlin B. The Indiana network for patient care: a working local health information infrastructure. Health Aff (Millwood). 2005 Sep-Oct;24(5):1214–20.

[10] Frisse ME, Tang L, Belsito A, Overhage JM. Development and use of a medication history service associated with a health information exchange: architecture and preliminary findings. AMIA Annu Symp Proc. 2010 Nov 13;2010:242-5.

[11] Sikka R, Xia F, Aubert RE. Estimating medication persistency using administrative claims data. Am J Manag Care. 2005;11(7):449–57.

[12] Hansen RA, Kim MM, Song L, Tu W, Wu J, Murray MD. Comparison of methods to assess medication adherence and classify nonadherence. Ann Pharmacother. 2009 Mar;43(3):413-22.

[13] Hong JS, Kang HC. Relationship between oral antihyperglycemic medication adherence and hospitalization, mortality, and healthcare costs in adult ambulatory care patients with type 2 diabetes in South Korea. Med Care. 2011 Apr;49(4):378-84.

[14] Lau DT, Nau DP. Oral antihyperglycemic medication nonadherence and subsequent hospitalization among individuals with type 2 diabetes. Diabetes Care. 2004 Sep;27(9):2149-53.

[15] Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, Magid DJ.. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-41.

[16] Stratton IM, Adler AI, Neil HA Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ. 2000 Aug 12;321(7258):405-12.

[17] Schectman JM, Nadkarni MM, Voss JD. The association between diabetes metabolic control and drug adherence in an indigent population. Diabetes Care. 2002 Jun;25(6):1015-21.

[18] American Diabetes Association. Standards of medical care in diabetes 2010. Diabetes Care. 2010. 33(Supplement 1); S11-S61.

[19] Lynch J, Smith GD, Harper S, Hillemeier M. Is income inequality a determinant of population health? Part 2. U.S. National and regional trends in income inequality and age- and cause-specific mortality. Milbank Q. 2004;82(2):355-400.

Address for correspondence

Vivienne J. Zhu

viviennezh@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-64

Validation of Minimum Data of Archetyped Telehealth Clinical Report for Monitoring Prenatal Care

Danielle Santos Alves[a], Valéria Cesário Times[a], Magdala de Araújo Novaes[b]

[a] *Center for Informatics, Federal University of Pernambuco (CIn/UFPE), Pernambuco, Brazil*
[b] *Telehealth Center of Pernambuco, Federal University of Pernambuco (NUTES/ UFPE), Pernambuco, Brazil*

Abstract

Studies on the validation of minimum data sets from international information standards have drawn the attention of the academic community to the identification of necessary requirements for the development of Electronic Health Records (EHRs). The primary motivation of such studies is the development of systems using archetypes. The aim of this study was to validate the minimum data set that should be used when constructing an archetyped EHR for prenatal care applications in telehealth. In order to achieve this, a data validation tool was built and used by nine expert obstetricians. The statistical analysis employed was the percentage of agreement and the content validity index. The study was conducted in three steps: 1) Literature review, 2)Instrument development, and 3) Validation of the minimum data set. Of the 179 evaluated pieces of data, 157 of them were validated to be included in the archetyped record of the first prenatal consultation, while 56 of them were allocated for the subsequent consultation record. The benefit of this research is the standardization (data validation for an archetyped system) of prenatal care, with the perspective of employing, both nationally and internationally, an archtyped telehealth system.

Keywords:

Validation; Archetype; Electronic Health Record; Obstetrics; Prenatal Care.

Introduction

Validation studies can be focused on any area of health. Primary care, especially in the form of programs focused on womens' healthcare, has been drawing attention due to its great impact on society and its repercussion on important maternal and neonatal morbidity and mortality indicators [1].

The monitoring of this population occurs at the level of primary care, in other words, it is made by the family health units (FHU) of each county throughout the state. However, it is known that there are intrisic issues with primary healthcare provision that are preventing it from fulfilling its objective. There are not enough primary care specialists for the whole population, making it therefore necessary to integrate the primary health network with secondary and/or tertiary levels of care.

For this integration, the FHU's doctor and/or nurse need to have tertiary network support in case there is the need to provide prenatal monitoring of their high-risk pregnant women. Telehealth can provide the necessary support in a timely manner, avoiding unnecessary trips and referrals, and enabling the population to remotely access and depend on specialists.

To facilitate communication between primary and tertiary levels through telehealth, it is necessary to use archetyped telehealth systems in obstetrics to allow for the exchange of information between them, and thus, detect early complications, avoid unnecessary referrals, and reduce the physical and financial costs of trips.

Studies and recommendations made by the Brazilian Health Informatics Association concerning the validation of minimum data sets of international information standards have drawn the attention of the academic community to the development of an Electronic Health Record (EHR) [2].

Furthermore, the development of systems using archetypes has been recommended by researchers [3]. Also, it has been recognised that in validation studies, the participation of health professionals is crucial. Therefore, with the help of nine experts, all obstetricians, this work aims to validate the minimum data set that should be used in the construction of an archetyped EHR for prenatal care applications in telehealth.

Many countries, including Brazil, are encouraging the use and adoption of international standards in the EHR to facilitate the exchange of information between systems and among the various healthcare levels. These recommendations have been proposed since 2011 by a decree from the Ministry of Health. It regulates the use of interoperability standards and health information for EHRs of the national health systems in municipal, district, state, and federal levels [3].

Currently, the major difficulty in the development of EHRs is that systems are built without performing minimum data set validation studies, in addition to the absence of terminology and universal standards. All of those issues, in the future, could undermine a proper and safe exchange of information between systems (interoperability), harming patients' well-being [4].

Among the proposals to help reduce the complexity barriers of interoperability and adopting standards in health are the studies proposed by the openEHR Foundation, which is a network of researchers working throughout the world, developing and testing the applicability of archetypes in healthcare systems.

Archetypes may be conceptualized as a set of specifications that define a reference model of health information, i.e., a language for constructing "clinical models" [5]. Archetypes define the concepts used in the clinical field, such as a clinical summary of a pregnant woman or the information on the patient's weight. An international repository–built by the OpenEHR foundation–already exists, containing validated archetypes that are used by researchers, developers, or health professionals while developing an EHR for many areas of health. Some of these archetypes are classified as validated, while others are under construction or being reviewed. Regarding pieces of information on women's health,

especially those related to prenatal care, the archetypes available in the OpenEHR repository may be considered weak. These archetypes are under construction and the information does not fully include the data considered important for the prenatal care attendance. One can also notice issues in the information about prenatal care in different archetypes, including dichotomy and non-sequential logic in the presentation of data related to childbirth and newborn care.

Studies show that EHRs are better received and evaluated by domain experts if they go through a process of content validation involving domain experts [6]. Hence, when seeking a more reliable structure that could also provide a higher level of satisfaction when using the EHR, validation studies with domain experts are still needed.

Before building EHR archetyped systems, content validation is needed and comprises more than a simple evaluation by a committee of experts [7]. It actually involves a trial process that consists of three steps: 1) literature review, 2) instrument development with the data, and 3) assessment of the minimum relevant data set to be used in a certain context by a committee of experts [8]. The main motivation of these studies is to develop systems using archetypes so that the women's pregnancy information, harvested during the prenatal care in a primary care setting, can be easily exchanged with specialists offering tertiary care through an archetyped clinical form of telehealth in obstetrics [3]. To achieve this, content validation studies are extremely necessary, especially if they are conducted with the participation of health professionals.

Thus, taking into account the importance of public health focused on women's health programs, and considering the national and international recommendations for structuring archetyped EHRs, the need for reviewing and validating the archetypes of the openEHR's database with the help of experts in obstetrics was identified. This process would enable faster construction of archetyped healthcare systems.

This study aims to validate the minimum data set that should be used in the construction of medical records of archetyped telehealth for prenatal care.

Methods

A descriptive and prospective study of quantitative and qualitative methodological types was conducted to determine and validate the minimum data set of a clinical record for prenatal monitoring in obstetrics [7]. The medical record must be a guiding resource for the health professional in comprehending the patient's history and physical examination, and during the registry of a care plan established for the customer. In the telehealth scenario, this clinical record must contain the minimum data required for the exchange of information between health professionals and experts. To provide a data standardization of this clinical record, metadata known as archetypes were used. This study was conducted among experts in obstetrics and expert teachers from the Federal University of Pernambuco (UFPE). In order to define the sample size, the literature recommends [8] a minimum of five and a maximum of ten people to take part in the study. Additionally, the training, qualification, and availability of the necessary professionals should be taken into account. Having observed these criteria, our sample consisted of nine obstetricians who agreed to participate in this research by signing a free and informed consent term. A methodological triangulation method [9] was used, where at least two kinds of methods (i.e. a quantitative approach and a qualitative approach) were adopted. For better understanding, the methodological steps were divided into 3 tasks: 1) literature review and identification of validated instruments to monitor/evaluate prenatal care based on standards of the Brazilian Ministry of Health and those of the World Health Organization (WHO), 2) preparation of the minimum data set, which was grouped into a single instrument using the concept of archetypes for structuring the clinical record of telehealth for prenatal care, and 3) validation of clinical records with experts. The reviewing process made by specialists was initiated with the invitation of members. These were contacted by telephone and/or email, and the purpose of the study was explained to them.

Upon the acceptance to participate, an online questionnaire with medical records' data was sent for validation. A period of 15 days was given to assess/respond to the questionnaire. The evaluation was made by selecting/classifying data. These were arranged in groups according to the exams' clinical approach, such as "history of the present pregnancy" and "general physical and obstetric examination."

The experts had to judge the degree of relevance of each data item included in the first consultation record and in the subsequent consultation record. In other words, the experts had to determine whether the questionnaire's data items should remain or not in the clinical record.

This project has been approved by CAAE: 33667214.3.0000.5208, according to the resolution number 466/12 of the National Health Council in Brazil.

Data Collection, Storage And Analysis

Data collection was made by filling out an online questionnaire with the pre-selected data items for validation. As indicated in the literature [10], a brief socio-demographic characterization of these experts was included in the questionnaire alongside a list of aspects related to pregnancy monitoring. The collected data was automatically entered or transcribed into a Microsoft Excel® spreadsheet, 2010 version. For the statistical purposes of this study, the concordance percentage calculation metrics (percent concordance) and the content validity index (CVI) were used. The percent concordance method was used to calculate the percentage of concordance between judges. It is the simplest measure of interobserver agreement [7].

The formula used is described below: percent concordance = (number of participants who agreed / total number of participants) x100 [7]. For this study, we considered 80% as an acceptable rate of agreement among the evaluators [7].

The content validity index (CVI) comprises a widely used method in health [7]. It measures the proportion or percentage of judges who are in agreement about certain aspects of the instrument and its data items. This method uses a Likert scale with scores ranging from one to five.

Experts are to qualify the degree of relevance of the qustionnaire's data item being evaluated by judging them on a Likert scale of 1 to 5, classifying them into 1 "Totally Irrelevant," 2 "Irrelevant," 3 "Indifferent," 4 "Relevant," and 5 "Totally Relevant."

The formula to calculate each individual item is: IVC = number of "4" or "5" responses / total number of responses.

To evaluate the instrument as a whole, there is no consensus in the literature. Polit and Beck [8] argue that in the evaluation process of data items individually, one should consider the number of judges. In the case of six or more, a rate of no less than 0.78 [7,8] is recommeded. To verify the validity of new instruments in general, some authors have suggested a minimum agreement of 0.80 [11,12]. The data items considered "approved" were those which obtained marks above 0.78.

Results

We interviewed nine experts. With respect to socio-demographic variables to characterize the experts, five respondents were teachers from UFPE and four were obstetricians who are currently working in delivery rooms of many endorsed health institutions located in the state of Pernambuco, Brazil. The average training time was 11 years and 5.3 years of practice. The average age of respondents was 32.4 years. Regarding the titles, two were PhDs, three had Master's degrees and four were experts only. Moreover, among the respondents, five participated in a research group in obstetrics.

Validation Of Minimum Data For The Archetypes

Of the 179 evaluated pieces of data, a total of 157 items were validated to be present in archetyped records of the first prenatal consultation, while 56 data items were considered important for inclusion in the subsequent consultation record. Based on the rates of percent concordance and CVI, Table 1 shows the final validation of minimum data for the construction of archetyped clinical records of prenatal care applications in telehealth. This table lists the categories used to group the data items investigated, the total data set that was analyzed and validated, the average percent concordance for the first consultation and for the subsequent consultation and, finally, the average CVI for the group as a whole. A detailed table with the validation information of each data item can be found at: https://docs.google.com/spreadsheets/d/1YXbSXcxARR8mlewR22vUCXMAaIoPezu-PTWgt1hcxaI/pubhtml

The variables of Category 1 concerned data identification. For the first consultation, all data items validated by experts were seen as relevant, except the data item related to the SISPRENTAL code. Experts suggested adding it to the subsequent clinical record.

For Category 2, which is related to socio-demographic data, all of the data items were validated for the first consultation and no data items were relocated to the subsequent consultation. Categories 3 and 4 concerned personal and family backgrounds. For these categories, all data was considered relevant to the first consultation document, except the variables "infection with hepatitis B / C" and "infection with syphilis" from Category 4. Also, no data items were considered to fit better in the subsequent clinical record.

For Category 5, 40 data items related to gynecological and obstetric backgrounds were analyzed. Data items concerning menarche, history related to sexual practices, and even data from previous pregnancies, including the number of births and the obstetric route, were evaluated. For the first visit, 38 items were considered relevant. For the subsequent consultation, the opposite considerations were made; all data items were evaluated as being irrelevant to the clinical records, except those related to the presence, type, and treatment of sexually transmitted infections. Data items related to the number of days of the menstrual cycle and the frequency of sexual activity were considered irrelevant to be present in the two types of prenatal care records.

Category 6 contained data about the current pregnancy: date of last menstrual period (LMP), expected delivery date (EDD), calculation of gestational age by LMP, ultrasounds of the 1st quarter, vaccination schedule for hepatitis and tetanus, admissions during pregnancy, and date of the beginning of prenatal care. All of these data items were considered relevant to be in the record of the first visit, except the information about previous admissions, the admission reason, and the gestational age (GA) of early prenatal care, which were considered irrelevant.

In Category 7, data related to the physical examination was included. All of the information was considered relevant to be in the first and subsequent records, and, in fact, the values of the CVI rates collected for them were equal to 1.00. In the category of obstetric physical examination (Category 8), only seven variables, out of the 15 data items studied, were considered relevant for the first consultation, albeit all of them were considered crucial to the subsequent consultation record.

With respect to the general physical examination (Category 9), despite all information having been considered relevant to the first consultation record, none of them were deemed relevant in the subsequent visits. Regarding clinical examination (Category 10), only five elements remained associated with the first visit record, and only three of them were added to the subsequent clinical record.

Table 1 – Final validation of minimum data for archetyped telehealth clinical record for the prenatal consultation based on the percent of agreement and CVI. Recife, 2014.

Category of validated data	Analyzed data	% of concordance (1st consultation)	% of concordance (subsequent Consultation)	Average of Likert scale
1 – Data Identification	8	0.94	1.00	4.72
2 – Sociodemographic Data	10	1.00	-	4.76
3 – Clinical Personal Background	27	0.99	-	4.76
4 – Family Background	13	0.94	-	4.62
5 - Gynecological and Obstetric History	40	0.97	1.00	4.70
6 - Pregnancy History	10	0.99	1.00	4.80
7 - Physical Examination - Vital signs, weight and height	6	1.00	1.00	5.00
8- Obstetric Physical Examination	15	1.00	0.95	4.53
9 - General Physical Examination	10	1.00	-	4.62
10 – Clinical Examination	7	1.00	1.00	4.43
11 – Images and Laboratory Wxamination	23	1.00	1.00	4.62
12 – Obstetric Guidelines	10	1.00	1.00	5.00

In Category 11, which corresponds to laboratory and imaging tests, only 15 variables, from the 23 analyzed pieces of data, were included in the first consultation record, and 18 variables remained in the subsequent consultation.

In the category of obstetric guidelines (Category 12), the guidelines evaluated were related to several aspects, namely: foods, discomforts of pregnancy, feeding, delivery, vaccination, laboratory tests, pap smears, return to subsequent consultations, use of ferrous sulfate, and use of folic acid. For the first consultation, the guidelines related to childbirth and lactation were considered irrelevant, but, all of them were considered important to the subsequent consultation. All categories reviewed and approved by the experts produced VCI values greater than 0.78 and the overall average amounted to 4.7 points on the Likert scale.

Discussion

The validation of the data items of our questionnaire showed interesting results regarding the relevance and even the permanence of some variables that were previously seen as important, and many are present in the pregnancy archetypes of the openEHR's data repository. This validation may therefore give rise to some discussion regarding the actual importance of each data item studied.

Before the construction of any computer system, researchers have recognised both the importance and need for an investigation with human subjects to validate openEHR's archetyped records [13] referenced in the literature. However, there was a lack of studies with similar objectives to the work described here. With the active participation of users in the process of planning, implementation, and evaluation of computer systems, a better acceptability of such a system is achieved when it is inserted into the users' work process [13]. In the writing of this article, many gaps related to archetypes validation for the development of systems were found, especially in obstetrics in primary care, a subject on which no studies were found in international databases during the literature review performed.

Through the use of a telehealth archetyped system based on a prenatal follow-up protocol, the exchange of standardized information between systems may be increased (interoperability) and the satisfaction of health professionals with the use of these systems may be improved as well (usability).

Conclusion

Validation studies are important tools to determine which data items are relevant for EHRs according to health professionals in the scenario of caring for pregnant women.

Data validation studies are referred to as being very important in the implementation and acceptance process of EHRs by health professionals, since the validation reduces data redundancy, builds more attractive systems, and enables the management of data relevant to health care applications, hence promoting better usability.

The results of the research described here allowed for the identification of essential data from the prenatal care attendence of pregnant women in primary care, with the perspective of interaction and exchange of information with specialists (tertiary care) through telehealth.

From these results, the archetypes available in the openEHR's database will be analyzed, and comparative validation of these archetypes will be made. Moreover, an analysis of the need to review existing archetypes will be proposed.

In this sense, and based on the results of this research, the construction of an archetype that includes all the data validated for prenatal care will be designed, and this will be submitted to the openEHR's evaluation committee. Finally, as another indication of future work, there is the proposal for the construction of an archetyped clinical record for telehealth systems and the validation of its interface through usability metrics.

References

[1] Ministério da Saúde B. Pré-natal e Puerpério: atenção qualificada e humanizada – manual técnico. 3rd ed. Brasil, editor. Brasilía: Ministério da Saúde - Brasil; 2006. 1-163 p.

[2] SBIS SB de I em S. Manual de Certificação para Sistemas de Registro Eletrônico em Saúde (S-RES). 4.1 ed. Saúde SB de I em, Medicina CF de, editors. Sociedade Brasileira de Informática em Saúde; 2013. 1-91 p.

[3] Ministério da Saúde B. PORTARIA Nº 2.073, DE 31 DE AGOSTO DE 2011 - Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os [Internet]. Brasilía: Ministério da Saúde; 2011 [cited 2014 Nov 26]. Available from: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073_31_08_2011.html

[4] Kalra D. Barriers, approaches and research priorities for semantic interoperability in support of clinical care delivery. 2008.

[5] OpenEHR. openEHR [Internet]. 2014. Available from: www.openehr.org

[6] Alves DS, and Novaes M de A. Profile of the nursing team and level of satisfaction in relation to the usability of the electronic patient's record. J Nurs UFPE line 2013: 7(1): 143–52.

[7] Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. Cien Saude Colet 2011:16(7): 3061–8.

[8] Denise F. Polit, and Beck CT. The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. Res Nurs Heal [Internet]. 2006 [cited 2014 Dec 14];29:489–97. Available from: http://cfd.ntunhs.edu.tw/ezfiles/6/1006/attach/33/pta_6871_6791004_64131.pdf

[9] Risjord MW, Dunbar SB, and Moloney MF. A New Foundation for Methodological Triangulation. J Nurs Scholarsh [Internet]. 2002 Sep [cited 2014 Dec 14];34(3):269–75. Available from: http://doi.wiley.com/10.1111/j.1547-5069.2002.00269.x

[10] Rubio DM, Berg-Weger M, Tebb SS, Lee ES, and Rauch S. Objectifying content validity: conducting a content validity study in social work research. Soc Work Res [Internet]. 2003 Jun [cited 2014 Dec 14];27(2):94–104. Available from: http://swr.oxfordjournals.org/content/27/2/94.short

[11] Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res [Internet]. 1992 Nov [cited 2014 Dec 14];5(4):194–7. Available from: http://www.researchgate.net/publication/248867629_Instrument_review_Getting_the_most_from_a_panel_of_experts

[12] Grant JS, and Davis LL. Selection and use of content experts for instrument development. Res Nurs Health

[Internet]. 1997 Jun [cited 2014 Dec 14];20(3):269–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9179180

[13] Van Der Meijden MJ, Tange HJ, Troost J, and Hasman A. Determinants of success of inpatient clinical information systems: a literature review. J Am Med Inform Assoc [Internet]. ????? Jan [cited 2014 Dec 21];10(3):235–43. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=342046&tool=pmcentrez&rendertype=abstract

Address for correspondence

Center of Informatics. Jornalist Anibal Fernandes Avenue, s/n, University City, (Campus Recife), ZIP code 50.740-560 - Recife – PE. (+55 81) 2126.8430. dsa3@cin.ufpe.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-69

One Million Electrocardiograms of Primary Care Patients: A Descriptive Analysis

Emmanuel Chazard[a], Milena Soriano Marcolino[bc], Chloé Dumesnil[a], Alexandre Caron[a], Daniel Moore F. Palhares[c], Grégoire Ficheur[a], Barbara C. A. Marino[c], Maria Beatriz M. Alkmim[c], Régis Beuscart[a], Antonio Luiz Ribeiro[bc]

[a] *Department of Public Health, CHU Lille; EA 2694, University of Lille; Lille, France*
[b] *Medical School, Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil*
[c] *Telehealth center, University Hospital of the Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil*

Abstract

In 722 cities of Minas Gerais (Brazil), primary care patients can have their ECGs remotely interpreted by cardiologists of the Telehealth Network of Minas Gerais (TNMG), a public telehealth service. As of December 2014, more than 1.9 million ECGs were interpreted. This study analyzed the database of all ECGs performed by the TNMG on primary care patients from 2009 to 2013 (n=1,101,993). Structured patient data and the results of automated ECG interpretation by the Glasgow Program are described. Mean patient age is 51 years old, 59% of them are women. The average body mass index is 25.9 kg/m², with an average increase of 0.15 kg/m² per civil year. Those patients notably have hypertension (33.2%), family history of coronary artery disease (14.5%), smoking (6.9%), diabetes (5.8%), obesity (5.8%) or Chagas Disease (3.0%). Seventy percent of ECGs are normal. This percentage is higher in women (72.3%) and decreases in average by 7.4 every 10 years of life. There are notably 12% of possible myocardial infarction, 10% of possible left ventricular hypertrophy and 8% of possible supraventricular extra systole.

Keywords:

Telemedicine; Electrocardiography; Cardiology; Big Data; Primary Healthcare.

Introduction

The physicians' diagnostic process in primary care is deeply impacted by the representation they form about the prevalence rates of the diseases (pre-test probabilities). Indeed, general practice "has a specific decision making process determined by the prevalence and incidence of illness in the community" [1]. Yet, this decision making process might be biased by the fact that most scientific studies focus on specific populations, especially hospital inpatients, that have increased prevalence rates of severe diseases. To support the diagnostic process in primary care, it is important to publish and discuss descriptive analysis of patients from primary care databases.

The electrocardiogram (ECG) is a widely available method that enables evaluation of the cardiovascular system, to diagnose some diseases or to exclude them. ECGs are easy to perform and available at low-cost. However, their interpretation is not that simple, and in many situations, the general practitioner (or his/her assistant) can perform the exam but is not able to interpret it. For that reason, it has been proposed to have the ECGs interpreted remotely by cardiologists [2]. The first experiment was conducted in 1905, only two years after the first electrocardiograph was ready for use.

To insure access to specialized healthcare to the Brazilian population of remote cities, the State Government of Minas Gerais (Brazil) funded in 2005 the Telehealth Network of Minas Gerais (TNMG) [3]. In this framework, a telecardiology program [4] enables the primary care physicians of remote areas of 722 cities to perform ECGs and transfer them in real time to an analysis center of the TNMG. The ECGs are analyzed by on-duty cardiologists, who send back free-text reports to the physicians the same day. Due to the high number of ECGs analysed per day (on average, 2,200 ECGs), there is a list of ECGs waiting for analysis. A nurse technician coordinates the distribution of the exams among the cardiologists on duty. Emergencies are analysed in less than 10 minutes and routine exams on average in 4 hours. The program has proven to be economically beneficial [5]. More than 1.9 million ECGs have been interpreted in December 2014 [6] but the analysis of this database has still not been published.

Only a few scientific papers provide descriptive analysis of large ECG databases, either as a main or secondary objective, and the results are informative to understand the patients' care and get reliable data about pre-test probabilities of heart diseases [7–9].

The objective of this paper is to provide the community with descriptive statistics about primary care patients and their ECGs in Brazil, by performing a descriptive analysis of the telecardiology database of the TNMG.

Methods

This retrospective observational study assessed the ECG database of the TNMG, which included all the consecutive exams performed from 2009 to 2013 (n=1,101,993).

The ECG database

The ECG database includes all exams performed in the remote health centers from 2009 to 2013, excluding emergency centers (0.2% of the sample). For each exam, the database contains:

- demographic information about the patient: gender, birth date, marital status, educational achievement, income;
- patient's symptoms: pain location (arms, neck, back, precordial, thoracic, epigastric), pain characteristics (caused by effort or emotion, relieved by rest or nitrates, pain intensity), other symptoms (dyspnea, sweating, vomiting, dizziness, syncope, palpitation);
- patient's clinical examination: height, weight, systolic and diastolic blood pressures;

- medications: diuretics, digitalis, beta-blockers, conversion enzyme inhibitors, amiodarone, calcium blockers, others;
- comorbidities or risk factors: arterial hypertension, obesity, diabetes, smoking, hyperlipidemia, personal history of myocardial infarction, personal history of coronary revascularization, family history of coronary disease, Chagas disease, chronic pulmonary disease, chronic kidney disease;
- ECGs represented by a list of time-dependent electric values available in 12 leads;
- administrative information about the ECG analysis: priority level, completion time, time of inclusion in the database, report writing start time, report sending time;
- a free-text report written by the on-duty cardiologist who interpreted remotely the ECG.

Automated computation of descriptive statements by the "Glasgow Program"

The Glasgow 12-lead ECG analysis program is computer software that can automatically interpret ECGs [10]. In this paper, it will be called "Glasgow Program," as in its technical documentation [11]. The Glasgow Program has been evaluated and has achieved very good results for signal processing, e.g., identifying waves and computing axes, durations, amplitudes and intervals [12]. It also obtained acceptable results for rhythm analysis and diagnostic interpretation [13,14]. A review paper concluded in 2001 that the output of computerized ECG processing could be used for epidemiological studies [15].

The Glasgow Program (release 28.4.1, issued on June 16th 2009) was fed with a structured description of the electric signal and a few clinical pieces of information, and was run on the total ECG dataset (n=1,101,993). It was able to output more than 900 different textual messages. We first defined a terminology according to the literature, in order to classify the ECGs the same way it is usually done in the scientific papers [8,9,16–22]. The ECG classification was performed in two steps. First, each ECG was classified into one of the following mutually-exclusive categories: normal, normal variant, abnormal, or fatal technical issue (the ECG cannot be interpreted). The "normal variant" category also contained ECGs that were normal except for rate. Secondly, only in case the ECG could be interpreted, the classification was made of several binary statements (diagnoses) that could be present (with different levels of sureness: "possible", "probable" or "certain") or not, independently from each other.

The Glasgow Program was also able to output quantitative descriptive variables, including heart rate; ventricular rate; average RR with standard deviation; heart rate variability; QRS, P, ST and T frontal axes; P, QRS, and T durations; and PR, QT and ST intervals. The corrected QT interval (QTc) was obtained using the Framingham correction. Those variables were analyzed only for "normal" ECGs, including normal variant.

Statistical analysis

Descriptive univariate statistics were computed from the whole database. For quantitative variables, the mean and standard deviation (SD) were computed. In case of non-normal distribution, quartiles were computed. A histogram was drawn, and/or a density line in case of continuous variable. For qualitative and binary variables, the proportions (prevalence rates) were computed and a bar plot was drawn.

The 95% confidence intervals of means were computed using the normal distribution when appropriate. The 95% confidence intervals of proportions were computed with the Exact Binomial Test, which can get reliable confidence intervals even when the smallest observed count is low or null [23]. The confidence intervals are not reported here when the sample size exceeds 200,000 because they are too narrow.

The Khi² test was used to test the independency between two categorical variables. Student t test or ANOVA was used to test the independency between a quantitative and a binary variable. Generalized linear model was used to test the independency between other kinds of variables. Moving average and linear regression were used to graphically represent the relation between a proportion and a quantitative variable. All the tests were double-sided and interpreted with a 5% significance threshold.

Data management and statistical computations were performed with R statistical computing software [24]. In this paper, when the p value of a test was lower than $2.2e^{-16}$, which is the computation limit of R, we simply wrote "p=0".

Variables having more than 25% missing values were excluded from the analysis. Missing values were studied.

Results

Background and clinical examination of the patients

The patients' mean age is 51 (SD=19.5; 0.5% of missing values). The distribution of the age is displayed on Figure 1. Females comprehended 59.3% of the sample, and this proportion is stable over the years. Patients' mean height is 1.61 m (SD=0.11; 23.7% of missing values). Patients' mean weight is 67.3 kg (SD=16.7; 21.4% of missing values). Patients' mean body mass index (BMI) is 25.9 kg/m² (SD=5.67; 23.5% of missing values; distribution displayed on Figure 2). The BMI is greater than 30 kg/m² in 20.3% of cases. The BMI values increase in average by 0.15 kg/m² per civil year (p=0).

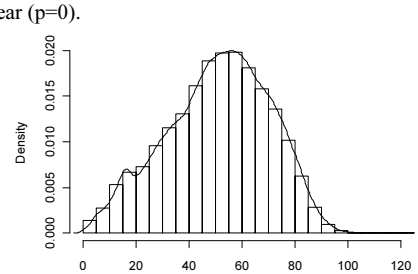


Figure 1 - Histogram of patient's age (years)

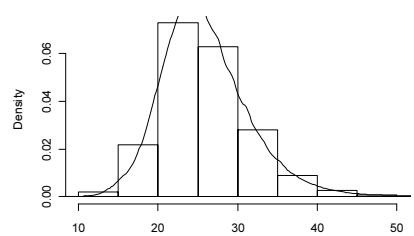


Figure 2 - Histogram of patient's body mass index (kg/m²)

Figure 3 displays the comorbidities and risk factors of the patients. Regarding the comorbidities, 33.2% of the patients have arterial hypertension, 14.5% have a family history of coronary disease, 6.9% are smokers, 5.8% have diabetes mellitus, 5.8% are obese, 3.1% have hyperlipidemia and 3.0%

have Chagas disease. Other statements have a frequency lower than 1%. However, 55.2% of patients are declared not to have any background or identified risk factor. As this part of the form is a set of checkboxes, it is not possible to know the proportion of missing values, but for instance 17.3% of patients having "no obesity" have a BMI greater than 30 kg/m².

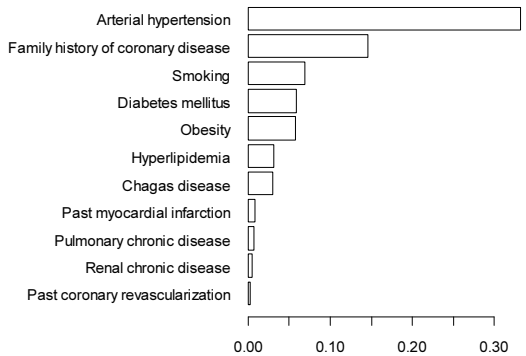


Figure 3 - Comorbidities and risk factors of the patients (prevalence rates)

Regarding the medications, 20.7% of patients take diuretics (the precise class is unknown), 7.9% beta-blockers, 2.8% digitalis, and 0.7% amiodarone. In addition, 95.7% of patients on diuretics suffer from arterial hypertension, and 59.5% of patients with arterial hypertension take diuretics. However, 68.6% of patients declared not to take any drug (see Figure 4). Another drug is taken in 11.0% of cases, but detailed information is only available as free text. As this part of the form is a set of checkboxes, it is not possible to evaluate the proportion of missing values.

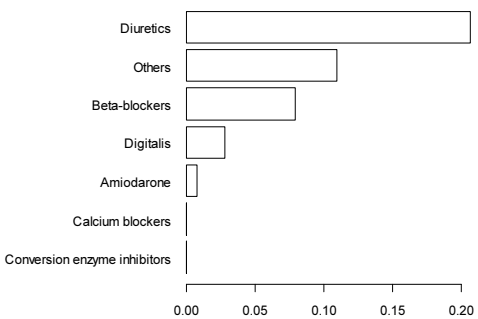


Figure 4 - Medications taken by the patients (prevalence rates)

Administrative management of the exams

At the time of the analysis, the database contains 1,101,993 different exams. The number of exams analyzed per year is displayed on Figure 5. As the year 2013 is incomplete at the time of the data extraction, the extrapolated annual value is represented using a dotted line.

Figure 6 displays the distribution of the time of the day the ECG was performed: 97.4% of them between 7:00 AM and 8:00 PM.

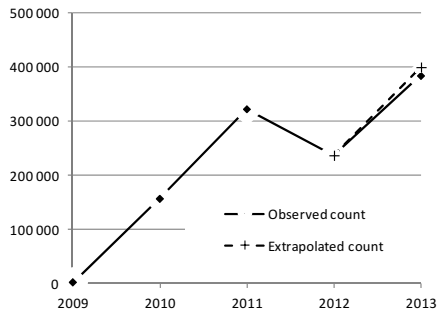


Figure 5 - Number of electrocardiograms performed per year

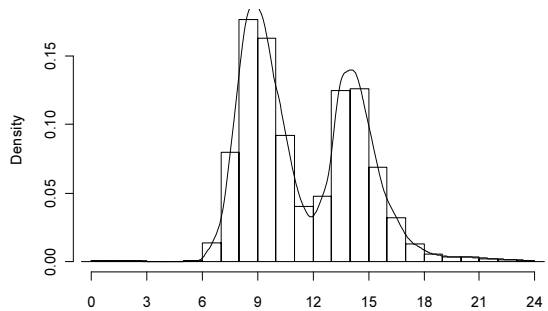


Figure 6 - Histogram of the time of the day the electrocardiogram was performed.

Prevalence rate of normal ECGs

After exclusion of fatal errors, the proportion of normal or normal variant ECGs is estimated to be 69.6%. The percentage of normal ECGs is significantly dependent on the patients' age (p=0), as illustrated on Figure 7. In average, it decreases by 7.43% every 10 years of patient life. The prevalence rate of normal ECGs is lower in men (65.8%) than in women (72.3%) (p=0). It appears to be stable over time (p=11.0%).

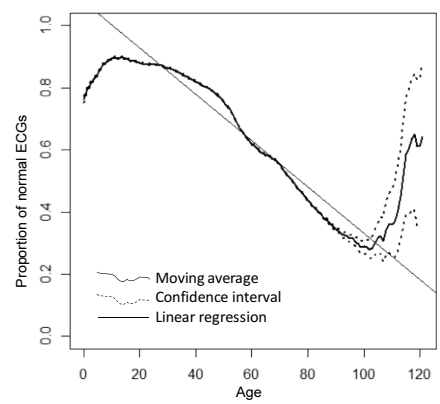


Figure 7 - Proportion of normal electrocardiograms (including normal variant) as a function of the patients' age

Electric description of normal ECGs

The electric parameters of normal ECGs (including "normal variant") are stratified according to gender, and detailed in Table 1. There are 0.00% to 0.86% of missing values for each measurement. The mean of every parameter is significantly different between men and women ($p<10^{-34}$ for each parameter).

Medical description of ECGs

The description of the final statements of the Glasgow Program is displayed in Table 2. The four statements are mutually exclusive. Then, except in case of fatal technical issue, the Glasgow Program can obtain detailed statements that are displayed in Table 3.

Table 1 - Mean (standard deviation) of automatically measured electric parameters of normal electrocardiograms

Variable	Women	Men	t test
Heart rate (BPM)	74.1 (14.8)	70.5 (16.0)	p=0
RR interval (ms)	826 (157)	876 (183)	p=0
PR interval (ms)	152 (28.0)	157 (30.5)	p=0
QRS axis (deg)	28.4 (37.5)	27.2 (44.8)	$p<1e^{-34}$
ST interval (ms)	109 (32.2)	94.0 (31.7)	p=0
QTc interval (ms)[a]	423 (22.8)	414 (24.1)	p=0

(a) with the Framingham correction

Table 2 - Final statement of the Glasgow Program

Final statement	Prevalence rate
Fatal technical issue	0.11%
Normal ECG	33.85%
Normal variant	35.78%
Abnormal ECG	30.25%

Table 3 - Prevalence rates of electrocardiographic statements ("possible", "probable" or "certain" statements)

Statement	Prevalence rate
Global statements	
Non-fatal technical issue	6.28%
Permanent pacemaker	0.45%
Sinus rhythm	93.24%
Non-exclusive ECG abnormalities	
Left ventricular hypertrophy	9.52%
Right ventricular hypertrophy	2.79%
Myocardial infarction	11.61%
Rhythm abnormalities	
Atrial fibrillation or flutter	3.59%
Multifocal or ectopic atrial rhythm	2.06%
Atrial or supraventricular extrasystole	8.21%
Sinusal bradycardia	4.98%
Sinusal or supraventricular tachycardia	3.62%
Accelerated or normal junctional rhythm	0.12%
Idioventricular rhythm	0.19%
Ventricular extrasystole	7.69%
Parasystole	0.00%
Ventricular tachycardia	0.08%
Conduction disturbances	
Sino-atrial block	0.02%
First degree atrioventricular block	5.78%
Second degree atrioventricular block	0.11%
Third degree atrioventricular block	0.15%
Wolff Parkinson White syndrome	0.68%
Left bundle branch block *	5.98%
Right bundle branch block *	4.15%
Descriptive ECG abnormalities	
Repolarization abnormality	39.36%
Bradycardia	5.29%
Tachycardia	3.55%
Short PR	1.99%
Low voltage	1.75%
QRS axis deviation	14.46%
Long QT	2.55%

**: complete, incomplete or fascicular*

Variables with too many missing values

Some variables could not be analyzed due to missing values. Demographic information about the patient included: marital status (48.0% missing), educational achievement (59.2% missing), income (68.8% missing). For those 3 variables, on average, the completeness percentage decreases by 15.4 per year (p=0).

Clinical examination of the patient: systolic and diastolic arterial pressures (respectively 89.1% and 88.7% of missing values). The completeness percentage increases by 1.2 per year (p=0).

Discussion

The objective of this paper was to analyze a million ECGs in a database and provide the community with up-to-date prevalence rates of cardiologic patient conditions in primary care population. This paper shows descriptions of the administrative management of the ECGs (number of exams, hours of performance), the patients (demographic variables, diseases, risk factors and drugs), the ECGs (medical diagnosis obtained from automated interpretation by the Glasgow Program, analysis of the proportion of normal ECGs). This information could participate in the knowledge that constitutes the base of decision-making in primary care.

The proportion of normal ECGs (around 70%) and its relation to the patients' age (it decreases by 7.4% every 10 years of life) are consistent with the literature [8]. The prevalence rates of the diseases are also consistent [8,9].

This study takes advantage of the huge database size that enables the researchers to get precise estimates of prevalence rates and means. However, the analysis raises the issue of completeness and reliability of clinical data. The last section of the results shows that many variables that may be considered secondary by the physicians have a lot of missing values. One can suppose that those values are not missing at random, and may be filled when they support the diagnostic, and left blank in other cases, but this cannot be verified. It is worth noting that the percentage of missing values increases year after year (about +15% per civil year for demographic variables), which probably illustrates the feeling for the physicians that their input is useless and time-consuming. However, this is not in contradiction with Brazilian guidelines for ECG interpretation: the ECG itself does not make the diagnosis. The cardiologist is in charge of interpreting the electric signal, but the clinician has to integrate the clinical symptoms, the risk factors, the personal history of the patient, and other exam results to make the final diagnosis. When the data are provided by the general practitioner, the absence of well shared definitions is probably harmful. In this database for instance, 17.3% of patients with "no obesity" have a BMI greater than 30 kg/m², and are then obese by definition.

The prevalence rates of electrocardiographic statements were obtained from an automated interpretation, and are not as reliable as expert analyses. Additionally, repeated exams from the same patients were not excluded.

Four distinctive characteristics of Brazilian patients can be observed in the analysis results. (1) Three percent of patients suffer from Chagas diseases. (2) Guidelines of the management of patients with arterial hypertension recommend thiazides as a possible first line treatment: 95.7% of patients under diuretics suffer from arterial hypertension, and 59.5% of patients with arterial hypertension receive diuretics (the precise class was not available). (3) The prevalence rate of obesity is high (the average BMI is 25.9 kg/m² and 20.3% of cases are over 30 kg/m²), and increases fast (the BMI values increase in average by 0.15 kg/m² per civil year). (4) Finally,

there are only a few smokers (6.9%), but this attribute could have been underreported.

Conclusion

This descriptive analysis of a huge ECG database brings informative results. The analysis may enable primary care physicians to take into account actual prevalence rates of patient conditions that play an important role in the process of medical decision making [1].

Acknowledgments

The authors would like to thank the funding organizations and programs: FAPEMIG (processes #EDT 2372/05 and PPM-00328-08), CNPq (process # 400934/2005-1), ARCUS, the State Department of Health of Minas Gerais (Brazil) and all collaborators from the university centers (UFJF, UFTM, UFU, UFMG, UNIMONTES) and towns involved in the Telehealth Network.

References

[1] The European Definition of General Practice / Family Medicine - Edition 2011 [Internet]. [cited 2014 Dec 18]. Available from: http://www.woncaeurope.org/content/european-definition-general-practice-family-medicine-edition-2011

[2] Hjelm NM, Julius HW. Centenary of tele-electrocardiography and telephonocardiography. J Telemed Telecare. 2005;11(7):336–8.

[3] Alkmim MB, Figueira RM, Marcolino MS, Cardoso CS, Pena de Abreu M, Cunha LR, et al. Improving patient access to specialized health care: the Telehealth Network of Minas Gerais, Brazil. Bull World Health Organ. 2012 May 1;90(5):373–8.

[4] Ribeiro ALP, Alkmim MB, Cardoso CS, Carvalho GGR, Caiaffa WT, Andrade MV, et al. Implementation of a telecardiology system in the state of Minas Gerais: the Minas Telecardio Project. Arq Bras Cardiol. 2010 Jul;95(1):70–8.

[5] Andrade MV, Maia AC, Cardoso CS, Alkmim MB, Ribeiro ALP. Cost-benefit of the telecardiology service in the state of Minas Gerais: Minas Telecardio Project. Arq Bras Cardiol. 2011 Oct;97(4):307–16.

[6] Centro de Telessaúde do Hospital das Clínicas da UFMG [Internet] [cited 2014 Dec 18]. Available from: http://www.telessaude.hc.ufmg.br/

[7] Park MY, Yoon D, Choi NK, Lee J, Lee K, Lim HS, et al. Construction of an open-access QT database for detecting the proarrhythmia potential of marketed drugs: ECG-ViEW. Clin Pharmacol Ther. 2012 Sep;92(3):393–6.

[8] Giuliano I de CB, Barcellos Junior CL, von Wangenheim A, Coutinho MSS de A. Issuing electrocardiographic reports remotely: experience of the telemedicine network of Santa Catarina. Arq Bras Cardiol. 2012 Nov;99(5):1023–30.

[9] Iacoviello L, Rago L, Costanzo S, Di Castelnuovo A, Zito F, Assanelli D, et al. The Moli-sani project: computerized ECG database in a population-based cohort study. J Electrocardiol. 2012 Dec;45(6):684–9.

[10] Macfarlane PW, Devine B, Latif S, McLaughlin S, Shoat DB, Watts MP. Methodology of ECG interpretation in the Glasgow program. Methods Inf Med. 1990 Sep;29(4):354–61.

[11] Physio-Control, Inc., Medtronic B.V. Statement of Validation and Accuracy for the Glasgow 12-Lead ECG Analysis Program [Internet]. 2009 [cited 2014 Dec 18]. Available from: http://www.physio-control.com/uploadedFiles/learning/clinical-topics/Glasgow%2012-lead%20ECG%20Analysis%20Program%20Statement%20of%20Validation%20and%20Accuracy%203302436.A.pdf

[12] Macfarlane PW, Devine B, Clark E. The university of Glasgow (Uni-G) ECG analysis program. Computers in Cardiology, 2005. 2005. p. 451–4.

[13] Salerno SM, Alguire PC, Waxman HS. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. Ann Intern Med. 2003 May 6;138(9):751–60.

[14] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med. 1991 Dec 19;325(25):1767–73.

[15] Kors JA, van Herpen G. The coming of age of computerized ECG processing: can it replace the cardiologist in epidemiological studies and clinical trials? Stud Health Technol Inform. 2001;84(Pt 2):1161–7.

[16] Kumar N, Saini D, Froelicher V. A gender-based analysis of high school athletes using computerized electrocardiogram measurements. PloS One. 2013;8(1):e53365.

[17] Hamilton RM, Houston AB, McLeod K, Macfarlane PW. Evaluation of pediatric electrocardiogram diagnosis of ventricular hypertrophy by computer program compared with cardiologists. Pediatr Cardiol. 2005 Aug;26(4):373–8.

[18] Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Zywietz C. Common standards for quantitative electrocardiography: goals and main results. CSE Working Party. Methods Inf Med. 1990 Sep;29(4):263–71.

[19] Widman LE, Tong DA. Validation of the EINTHOVEN model-based computerized electrocardiogram rhythm analysis system with three classes of clinical arrhythmias. Am J Cardiol. 1996 Oct 15;78(8):927–31.

[20] Guglin ME, Thatai D. Common errors in computer electrocardiogram interpretation. Int J Cardiol. 2006 Jan 13;106(2):232–7.

[21] Chiu CC, Hamilton RM, Gow RM, Kirsh JA, McCrindle BW. Evaluation of computerized interpretation of the pediatric electrocardiogram. J Electrocardiol. 2007 Apr;40(2):139–43.

[22] Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. J Electrocardiol. 2007 Oct;40(5):385–90.

[23] Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika. 1934 Dec 1;26(4):404–13.

[24] Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2009. Available from: http://www.R-project.org

Address for correspondence

Emmanuel Chazard, Public Health Department CHRU Lille, 2 avenue Oscar Lambret, 59037 Lille Cedex, France.
Email: emmanuel.chazard@univ-lille2.fr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-74

ePoint.telemed –An Open Web-based Platform for Home Monitoring of Patients with Chronic Heart Failure

Halgeir Holthe[a] , J. Artur Serrano[b]

[a] *NST-Norwegian Center for Integrated Care and Telemedicine, University Hospital North Norway, Tromsø, Norway*
[b] *NST-Norwegian Center for Integrated care and Telemedicine, University Hospital North Norway, Tromsø, Norway, Department of Clinical Medicine, Faculty of health Science, The Arctic University of Tromsø, Tromsø, Norway*

Abstract

In North Norway, no telemonitoring services for chronic heart failure (CHF) have yet been established, hence no investigations in the area have been published. However, large distances and a sparse population are causes for extra expenditure on hospital visits. In this paper, we describe the ePoint.telemed platform for home telemonitoring of CHF patients. We have reviewed the literature on home monitoring techniques, and developed two prototype platforms for remote collection of physiological data. We have refined one of the prototypes and subjected it to user testing among health professionals and their clients. Fifty patients will be involved in a randomized controlled trial aiming to establish if the home telemonitoring of CHF is clinically feasible and cost-effective. The ePoint.telemed platform is a fully automated internet based system meant for early warnings in a CHF rehabilitation program. The core of the platform is a dashboard connected to a blood pressure meter, a weight scale, and a web-based patient questionnaire. Unlike traditional systems built on dedicated medical equipment, we are applying easy-to-use components geared towards the sports market.

Keywords:

Chronic Heart Failure; Home Telemonitoring; Remote Monitoring Technologies; Personal Health Information Systems; Personal Health System

Introduction

In Norway, the costs for treatment of chronic heart failure (CHF) are huge, both in terms of hospital treatment and daily use of medication over years, as well as in terms of loss of quality of life for patients and their family caregivers [1]. Structured telephone systems and remote monitoring technologies (RMT) for monitoring CHF patients at home have been widely deployed in Europe over the last 20 years or so [2-7]. These systems are considered capable of increasing quality of life for the patients using them, to lower death rates, to reduce costs for medical treatment, and to decrease cases of re-hospitalization after heart surgery. However, home monitoring of patients with CHF is still in its beginnings in Norway, and little practical experience has so far been gained. In fact, to our knowledge, no routine telemonitoring services are available in Norway for C and hence no investigations in the field have yet been published, although large distances and a sparse population in N. Norway are a cause for extra public and private expenditure on travel for patients when making a visit to the hospital. These state of affairs constitute a strong motivation for home monitoring of CHF patients in Norway. But it seems that current evidence of the effectiveness and quality of CHF home monitoring systems is insufficient to recommend usage. The structure and funding streams in Norwegian health services are different from many other European countries and the outcomes of comparisons between conventional CHF services and the home monitoring systems in previous studies in Europe, may not fit well in a Norwegian context. It is thus important to further investigate the feasibility of home monitoring of patients with CHF in Norway. This paper describes a technological platform ePoint.telemed for daily monitoring of patients' weight, blood pressure, pulse, and self-declared health status directly from the patients' home. The system transmits the patient data automatically, seamlessly, and securely to a server at the University Hospital North-Norway (UNN), and allows examination of these values by a trained nurse at the hospital. An RMT system for monitoring CHF patients typically uses dedicated medical equipment to measure body weight, blood pressure, heart rate, blood oxygen saturation levels, and ECG. These systems are comparatively work intensive in terms of set-up and patient training. An RMT system provides asynchronous communication between patients and health personnel, and such systems are also remotely configurable. Compared with web-based solutions, structured telephone support requires synchronous follow-up by health-professionals and telephone systems are not remotely configurable.

The ePoint.telemed platform is a net-based system that works asynchronously with no need for interventions from health professionals, except on alarms. The platform is remotely configurable and may be individualized and set up without actually visiting the patients' home. This paper describes the technical details of the ePoint.telemed platform and the development process behind it. The platform is now being applied in a randomized trial at UNN to determine if the system is medically feasible and cost effective, if it is capable of reducing hospital re-admissions, and if it is acceptable to patients and their families as an integral part of a CHF rehabilitation program.

Methods

The prototyping of the ePoint.telemed platform was, from the start, driven by healthcare professionals at UNN. Medical doctors and nurses were involved in the design and implementation of the system from the prototyping stage until

the user trials. Moreover, early tests were performed with patients using prototypes. This may best be described as a combination of expert inspection and user-oriented design methods [8].

In 2011, we developed an alternative platform for home monitoring of CHF patients based on the dedicated medical equipment attached to a laptop in the patients' homes. The laptop was able to transmit data to our research laboratory servers using the patients' home internet connection. However, the interfacing of the medical equipment was far from trivial, and our practical experiments warned us that the maintenance of the system in the patients' homes was potentially expensive, requiring several man-hours to be set up properly. The collection of physiological data required individual adaptation of the software in each laptop due to individual needs among the patients. Since the setup of thresholds for specific biological parameters presumed programming on each laptop, our nurse had to review all of the received data manually. This was a quite laborious process and reconfiguration of our system was requested by the health professionals. We then decided to try out components geared towards the sports market for measuring physiological parameters.

Two different strategies for component integration are now implemented in the ePoint.telemed platform: the first one based on email parsing using the Representational State Transfer Application Programming Interface (REST API) [9] and a second one for pull of data based on the OAuth communication package [10]. Currently, three devices for data collection are integrated: a blood pressure meter, a weight scale, and a tablet computer. However, more and alternative types of components can easily be integrated into the platform due to its open design. At present, the platform integrates a Withings weight scale [11], facilitating push of data (weight measurements) via the easy-to-use Withings API. The BP3 blood pressure meter from iHealth [9] is also integrated, which delivers diastolic/systolic blood pressure and heart rate measurements. The tablet applied is the iPad, selected because of its compatibility with the iHealth system. Our platform is open and easily configurable for other types of equipment. The estimated effort for configuring a new device on the system is approximately one to two man-days.

All data transfer is anonymized by coding the patients' identities. In this project, confidentiality of private health information is ensured according to the Norwegian Ministry of Health and Care Service regulations [12]. All personal data which link measurements to a specific patient are removed and replaced by a code so that the information can be used alone or in combination with other health record entries. The relation between the codes and the patients' names are kept on paper in the heart polyclinic with a backup stored in a hospital PC not connected to any network.

We can collect vital physiological data according to medical needs, and set up user specific thresholds for any parameter in a fully automated fashion. The system will send mail-based alarms to users and/or health professionals for any data out of range as specified by the thresholds. In the pilot study, two patients with chronic heart failure were recruited through the Tromsø municipality home services. The patients receive their treatment at the University Hospital North-Norway, and participate in rehabilitation programs at the hospital. After being informed of the study, the patients gave written informed consent to participate in our study together with health professionals at UNN's heart polyclinic.

In 2013, after the pilot study, an application was sent to the Norwegian Regional Ethical Committee for ethical approval of a randomized controlled trial with 50 patients. After receiving the approval for the project, the final version of the ePoint.telemed platform was developed and the randomized controlled trial started in September 2014 at UNN. All patients admitted at UNNs' heart polyclinic, satisfying the inclusion criteria of the project, are randomized into an intervention arm testing the proposed telemonitoring service and a control arm with the treatment usually offered by the polyclinic. Patients assigned to the intervention arm receive a home telemonitoring kit consisting of a tablet, a wireless weight scale, and a portable blood pressure meter. All participants are given 30 minutes of instruction about the use of the equipment by a trained nurse. The patient is asked to weigh him or herself every day in the morning before breakfast. In addition, the patient also takes a measurement of the blood pressure and heart rate, and answers an online multiple-choice questionnaire about personal health conditions.

We have, through the advice from the health professionals in our project, chosen 10 questions derived from the Minnesota Living with Heart Failure Questionnaire (MLHFQ). The MLHFQ is a disease specific tool for measuring quality of life among patients living with CHF and the outcomes are according to T Hole co-releated to patients' level of functioning [13]. The MLHFQ can be divided into three categories of questions (with 8, 6, and 7 items respectively, each organized as entries on a 6 level score Likert scale).We concentrated on the questions about psychological status in the first two categories. Since the patients are expected to fill in the net-based questionnaire on a daily basis, it is necessary to keep it as short as possible. Our aim is, of course, not to assess the patients' quality of life as such but to obtain a fair estimate of any changes in psychological status. Patients assigned to the control arm receive conventional treatment.

Acceptance of the system is crucial for the successful implementation of a home monitoring service. Hence, we have included the aid of professionals as well as patients in our prototyping process. This means that the users have been involved in the project from the design of paper mock-ups and onwards. The patients participated in three design meetings together with health professionals and technologists before the installation of the prototype in the patients' home. After installation we have visited the patients at home routinely once a month, starting with instructing the patients about the use of the equipment. Later we have adjusted the system and checked the patients' operation of it. Since the fall of 2014, the system has been used on a daily basis. The cooperation with the health professionals (primarily the nurses at UNN's heart clinic) has taken the form of a continuous dialogue, involving several rounds of testing and adjustment of the running prototype. From our interactions with patients and professionals we can safely conclude that the prototype is accepted within the design team. However, a wider user acceptance of a home monitoring service presumes clinical validation and confirmed user friendliness in a larger cohort of patients.

Results

The prototype system deployed in this pilot study consists of a Withings wireless weight scale [11] and an iHealth blood pressure meter [9] that communicates with an iPad using the bluetooth protocol. Emails containing the blood pressure

measurements and the self declared health status of the patients are automatically generated by the system. These emails are then sent to a server at the hospital. The weight measurements are pulled automatically to a dashboard once per day and stored on the server. At the hospital side, the nurse receives an e-mail from the system every morning. The nurse actually has to check the incoming data only for alarms.

The dashboard may best be described as a service management system coded in XHTML and PHP. The dashboard allows for administration of system users, i.e. connecting patients and health professionals. The dashboard handles pull of weight measurements and extracts information from incoming e-posts. Individual thresholds for each biological parameter may be set. The data is stored in a MYSQL database management system. The weight measurements and the patients' health status declarations are available on the dashboard together with the blood pressure and heart rate measurements (Figure 1). The data are displayed graphically in an easy-to-read manner. After the first two months of pilot testing, the patients were satisfied with the usability of the solution. As for the hospital side, the data arrived in a timely and correct way, and the reliability of the system seems to be proven. The nurse also gave positive feedback on the system usability and functionalities.

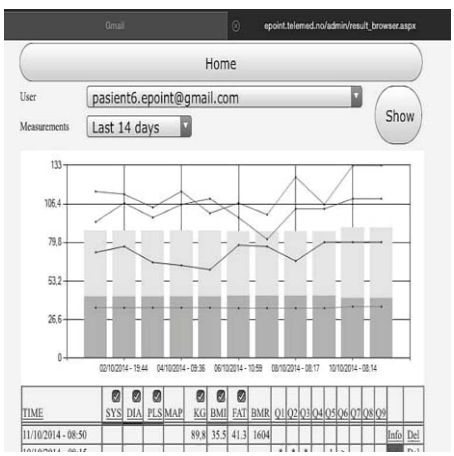


Figure 1-Snapshot of the Dashboard

The ePoint.telemed platform developed for the RCT is web-based and it follows an open strategy of component integration [14]. Patient data is accessible through a browser window. Health data are stored in a secure server. Only health professionals have access to the dashboard and the connected server. The patients can still read their data in gmail or in their Withings account. The ePoint.telemed platform applies easy-to-use Personal Health Systems (PHS) components geared towards the sports and wellness market [3]. The inbuilt internet connection of the components used in the platform implies a seamless communication with a central server placed in a local health service center or in a hospital. Vital data can be collected according to individual medical needs.

The measurements of both the blood pressure and heart rate are collected by the iHealth App on the iPad. The collected data is then sent through an email to the ePoint.telemed, back-end by a secured socket layer channel. Triggered at regular times (at the moment at 9am although these can be configured), a PHP script parses the CSV-file in the email using queries for insertion of the data into the MYSQL database. A REST API is used to parse the CSV-file. The parsed data collected by the iHealth, i.e. blood pressure and heart rate, is then presented in the dashboard.

The user is required to fill in a web-based questionnaire daily. This is accessed through a link icon in the iPad. Once the questionnaire is completed, an email is automatically composed with the answers. This is sent anonymously to the central dashboard as a CSV-file. The file is parsed in the same way as the process with iHealth, using the REST API. Weight measurements are pulled directly from the users' Withings accounts applying the OAuth communication package.

Discussion

The ePoint.telemed platform is a fully automatic internet based system meant for early warning in a CHF rehabilitation program. Unlike traditional RMT systems building on dedicated medical equipment, we are applying easy-to-use personal health systems (PHS) components geared towards the sports market. The components of the platform are fairly low cost and involve a minimum of installation efforts. The components automatically transmit the vital biological parameters to users' accounts at the vendors' internet sites. We pull the data from there, and log them into a dedicated database system. The low cost, the remote configurability, and the ease of installation and maintenance of our platform represent new features not present in conventional systems (including telephone based platforms) for home monitoring of patients with chronic heart disease.

Home monitoring of CHF potentially enhances the patients' situation during rehabilitation, and probably also means better service from the health professionals. Still, at present remote monitoring of vital symptoms is not a prioritized area in the treatment of chronic disease in Norway. For example, approximately 489 new pacemakers per million inhabitants are implanted annually in Norway (figures from 2009) [15]. But no systematic attempts have so far been made to make use of the data from these devices in CHF rehabilitation programs. However, home monitoring of COPD [16] is currently being performed in Norway, and in treatment of diabetes remote monitoring of biological parameters is at present strengthening self-management schemes as well as rehabilitation programs [17,18].

User oriented design methods have a strong position in Scandinavia, and we have emphasized including both patients and experienced health personnel in our prototyping process [19-21]. The successes of the ICT based health interventions are strongly dependent on user acceptance, both from the professional side as from the patient side. Usability issues in the ePoint.telemed platform were given careful attention. The platform is validated in a real clinical setting and follows the approved clinical procedures. The openness of the platform allows for the future inclusion of other devices (but the software used is proprietary). The platforms' applicability in connection with other pathologies such as diabetes or COPD can be supplementary tested with the inclusion of a glucose meter and a pulse oximeter, respectively.

The platform was then installed at the patients' home by a team from The Norwegian Center for Integrated Care and Telemedicine (NST). We used approximately 1 hour for installations and for instructing the patients about the basic operations of the equipment. All patients had a broadband network installed in their homes, prior to our trials. For home monitoring of heart patients in a Norwegian context, an easy-

to-install system with low maintenance costs is important. A rapid weight increase in connection with chronic heart disease (more than 1,5 kg in 2 days or more than 2,2 kg within a week) may suggest an accumulation of liquid in the body as a result of heart dysfunctions. Also, a significant and persistent weight loss over a few weeks may serve as a medical indication of problems in the heart function. We are in the process of confirming the accuracy of the collected data, but since our interest is primarily to detect changes in the parameters measured over time, we expect the precision of our data to be satisfactory.

On the hospital side, the nurse participating in the pilot study receives the patients' measurements each morning. During the pilot study, these measurements were inspected manually (in addition to the system intrinsic routines for measurement checking) and compared with baseline measurements taken at the start of the pilot study. The procedure for the patient was to turn on the weight scale and step on it, and then place the blood pressure meter on the arm to take a measurement. The patients were instructed to take a weight and blood pressure measurement every morning at approximately the same time, after visiting the toilet but before breakfast. Blood pressure values exceeding 150/110 mmHg was considered dysfunctional (but the health professionals can specify any threshold for this parameter after individual inspection of any patient).

After the first two months of pilot study testing and error corrections, the ePoint.telemed platform was considered operational both by the health professionals and the patients. As for the hospital side, the data arrived timely and correctly and the reliability of the system was accepted by the developers. Since the pilot study, we have further developed the user interface on the dashboard, and we have replaced the PC at the users' side with an iPad.

The primary outcome of the study is of administrative interest. The CHF related hospital re-admissions (calculated as the ratio between re-admissions in the intervention group compared with the re-admissions in the control arm) will be the key outcome of the study. We expect that early warning about any change in vital parameters may potentially reduce costs for interventions directed towards specific patients. Also, the economic benefit (additional costs divided by savings within the intervention group) resulting from applying the ePoint.telemed platform is of importance even if we plan to recruit only 50 patients. For the patients themselves, home monitoring of important biological parameters may imply an increased sense of security, and it may mean a decreased number of face-to-face meetings with the doctor during rehabilitation (saving time and travelling costs). In other words, we will investigate the economical benefits from applying the ePoint.telemed platform identified through a cost analysis between the intervention and control arms of the trial, and we will inquire the patients' own sense of wellbeing during rehabilitation. Measurements including weight, heart rate, blood pressure, and answers to the MLHFQ are taken at baseline, after 3 months and at study end (six months using the system).

In the practice field of healthcare, home monitoring of physiological data implies an increased focus on patients' self-management. The outcome may be reduced mortality rates for specific symptom categories and reduced costs per patient within specific rehabilitation programs [22, 23]. For the patients themselves it may mean a greater awareness about personal health and possibly an increased sense of security during rehabilitation. However, the implementation of routine home monitoring services at a heart clinic presumes organizational adaption and formalization of new working procedures among health professionals [24]. Hence, careful planning of the transitional stages is of vital importance when introducing new services in a clinical setting. Technically, more standardization of equipment and better integration with current health record systems seems to be necessary [25] to make home monitoring of CHF a common practice in Norway.

Conclusion

A platform for the home telemonitoring of patients suffering from chronic heart failure has been designed and implemented at the UNN. A user-oriented design method was applied involving both health professionals and patients. The prototype has been refined into a test version in 2014. The current version is being used in a real setting in the context of a randomized controlled trial. Approximately 50 patients will be participating in the study. The platform designated as ePoint.telemed presents innovative features that in their combination are unique to home monitoring of patients with chronic heart failure.

These features include:

- A fully anonymous transmission of data
- An open design of the platform allowing for the integration of devices from different suppliers
- A two stage patient questionnaire designed in cooperation with a hospital polyclinic
- A fully configurable set of individualized patient thresholds
- Automatic feedback to patients via e-mail
- Full user access to the clinical data collected
- A web-based dashboard with secure access for health professionals.

The success of ICT based health interventions is strongly dependent on user acceptance, both from the professional side and from the patients' side. Usability issues in the ePoint.telemed platform were given careful attention. The combinations of features of the ePoint.telemed Platform meet three fundamental requirements of electronic medical equipment in Norway: secure data transmission and storage, clinical efficacy, and a high degree of usability both for patients and professionals. The functionality of the platform was validated during real clinical testing.

Acknowledgements

We thank Helse Nord for the funding and Ronny Broekx, Ragnhild Jonassen and Marina Nilsen for their contribution during prototyping and implementation of the system.

References

[1] Norwegian Cardiovascular Register. Report from workgroup appointed by Norwegian Institute of Public Health. Oslo:2008. Available from: http://www.regjeringen.no/upload/HOD/Dokumenter%20FHA/H%C3%B8ring%20om%20hjerte-%20karregister200900661/Hjerte-karrapport_2008-10-03_%202009-02-16.pdf

[2] Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, Steward S and Cleland JG. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010; 8:CD007228.

[3] Baum P, Abadie F, Lupiañez FV, Maghiros I, Mora EV, and Talaya MBZ. Market Developments–Remote Patient Monitoring and Treatment, Telecare, Fitness/Wellness and mHealth. 2013. Available from: http://ftp.jrc.es/EURdoc/JRC71141.pdf

[4] Tunstall.[cited 23.october. 2014]. Available from: http://uk.tunstall.com/

[5] Winkler S, Schieber M, Lücke S, Heinze P, Schweizer T, Wegertseder D, Scherf M, Nettlau H,Henke Sascha and Braecklein M.A new telemonitoring system intended for chronic heart failure patients using mobile telephone technology-Feasibility study. Int J of Cardiol 2011; 153(1): 55-58.

[6] Dinesen B, Haesum LK, Soerensen N,Nielsen C, Grann O, Hejlesen O,Toft E, and Ehlers L. Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients. J Telemed Telecare 2012; 18(4): 221-225.

[7] Pinna GD, Maestri R, Andrews D, Witkowski T, Capomolla S, Scanferlato JL, Gobbi E, Ferrari M, Ponikowski P, Sleight P, Mortara A and Johnson P.Home telemonitoring of vital signs and cardiorespiratory signals in heart failure patients: system architecture and feasibility of the HHH model. Int J of cardio 2007; 120(3): 371-379.

[8] Bastien JM. Usability testing: A review of some methodological and technical aspects of the method. Int J Med Inf 2010; 79 (4): 18-23.

[9] ihealth Labs. [cited 25.october.2014].Available from: http://www.ihealthlabs.com/

[10]OAuth. OAuth 2.0 authorization framework. [cited 25.october.2014].Available from: http://oauth.net/

[11]Withings. [cited 25. October. 2014]. Available from: http://www.withings.com/eu/

[12]Ministry of Health and Care Services:Health Personnel Act § 21. 2001. [cited 10.November.2014].Available from:https://lovdata.no/dokument/NL/lov/1999-07-02-64/KAPITTEL_5#KAPITTEL_5.

[13]Hole T, Grundtvig M, Gullestad L, Flønæs B, and Westheim A. Improved quality of life in Norwegian heart failure patients after follow-up in outpatient heart failure clinics: results from the Norwegian heart failure registry. Eur J Heart Fail 2010; 12(11): 1247-1252.

[14]Fielding RT.Architectural Styles and the Design of Network-based Software Architectures [PhD Dissertation]. Irvine, CA: University of California;2000.

[15]Haug B, Kjelsberg K, and Lappegård KT. Pacemaker implantation in small hospitals:complication rates comparable to larger centres. Europace 2011; 13(11):1580-6.

[16]Zanaboni P, Lien LA, Hjalmarsen A, and Wotton R. Long-term telerehabilitation of COPD patients in their homes: interim results from a pilot study in Northern Norway. J Telemed Telecare 2013; 19(7): 425-429.

[17]Chomutare T, Fernandez-Luque L, Årsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011; 13 (3):e65.

[18]Helge FD, Årsand E, and Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012; 14 (4):e111.

[19]Kaulio MA. Customer, consumer and user involvement in product development: A framework and a review of selected methods. Total Qual Manage 1998; 9(1): 141-149

[20]Veryzer RW and Mozota BB. The Impact of User-Oriented Design on New Product Development: An Examination of Fundamental Relationships. Journal of Prod Innovat Manag 2005; 22(2): 128-143.

[21]Mao JY, Vredenburg K, Smith PW and Carey T. User-centered design methods in practice: A survey of the state of the art. In Proceedings of the 2001 conference of the Centre for Advanced Studies on Collaborative research. IBM Press, 2001; 12.

[22]Polisena J, Tran k, Cimon k, Hutton B, McGill S, Palmer K, and Scott RE. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010; 16(2): 68-76.

[23]Clark R A, Inglis SC, McAlister FA, Cleland JGF, and Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis.Bmj 2007.

[24]de Vries AE, van der Wal MH, Nieuwenhuis MM, de Jong RM, van Dijk RB, Jaarsma T, Hilege HL. Health professionals' expectations versus experiences of internet-based telemonitoring: survey among heart failure clinics. J Med Internet Res 2013; 15(1):e4.

[25]Pantelopoulos A, and Bourbakis NG. A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions Syst Man Cybern 2010; 40(1): 1-12.

Address for correspondence

Halgeir Holthe, Norwegian Center for Integrated Care and Telemedicine, University Hospital North-Norway, P.o.Box 35, 9038 Tromsø, Norway. Telephone:+4795729763.
halgeir.holthe@telemed.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-79

Mobile Healthcare System for Health Checkups and Telemedicine in Post-Disaster Situations

Min Hu[a,b,c], Megumi Sugimoto[b], Andrew Rebeiro Hargrave[b], Yasunobu Nohara[a], Michiko Moriyama[d], Ashir Ahmed[b,e], Shuji Shimizu[c], Naoki Nakashima[a,b,c]

[a] *Medical Information Center, Kyushu University Hospital, Japan*
[b] *Institute of Decision Science for a Sustainable Society, Kyushu University, Japan*
[c] *Telemedicine Development Center of Asia, Kyushu University Hospital, Japan*
[d] *Clinical Nursing Research Laboratory, Hiroshima University, Japan*
[e] *Department of Advanced Information Technology, Kyushu University, Japan*

Abstract

Portable Healthcare Clinic (PHC) is a mobile healthcare system comprising of medical sensors and health assessment criteria. It has been applied in Bangladesh for the last two years as a pilot program to identify non-communicable diseases. In this study, we adapted PHC to fit post-disaster conditions. The PHC health assessment criteria are redesigned to deal with emergency cases and healthcare worker insufficiency. A new algorithm makes an initial assessment of age, symptoms, and whether the person is seeing a doctor. These changes will make the turn-around time shorter and will enable reaching the most affected patients better. We tested the operability and turn-around time of the adapted system at the debris flow disaster shelters in Hiroshima, Japan. Changing the PHC health assessment criteria and other solutions such as a list of medicine preparation makes the PHC system switch into an emergency mode more smoothly following a natural disaster.

Keywords:

Paper, Medinfo 2015; Post Disaster; eHealth; mHealth; Teleconsultation; Public Health Informatics.

Introduction

The World Health Organization declared that 58 crisis countries, including Bangladesh, face acute Human Resources for Health (HRH) crisis [1]. Many countries experience catastrophic natural disasters. In Bangladesh, 26% of the population is affected by cyclones, and 70% live in flood-prone regions [2]. Natural disasters result in casualties and damage medical facilities and the health workforce [3]. The 2011 Tohoku Earthquake demonstrated the need for disaster management in aspects including healthcare. Although most casualties in the Tohoku Earthquake were because of drowning, the bad living environment in post-disaster shelters worsened the health condition of victims [4]. Strategies for post-disaster management are urgently required, and mobile healthcare systems may reach disaster victims quickly.

Kyushu University Hospital and Grameen Communications conducted a health management study using information communication technology. The Portable Healthcare Clinic (PHC) with medical sensors provided immediate consultation with the remote doctor over Skype for non-communicable diseases (NCDs). Following consultation, the remote doctor gives the patient an e-prescription. Data collection was conducted in 10 locations of Bangladesh between July 2012 and February 2014 [5].

However, little is known about the feasibility of introducing PHC into post-disaster areas where the healthcare conditions are different from those in non-disaster areas.

We examine the feasibility and risks of post-disaster healthcare management with a general health evaluation targeting disaster-related symptoms caused by trauma, infectious and chronic diseases, and mental disorders. After disasters, the findings of this study would be useful for developing an emergency mode of PHC to support post-disaster areas in Bangladesh, Japan, and other countries.

Methods

Logistics Classification Research

We collected data and performed a literature review on disasters in Bangladesh to understand the risks in post-disaster areas. We collected data from EM-DAT and Asian Disaster Reduction Center and used the keywords "natural disasters," "disease," "healthcare," "impact," "epidemiology," and "shelter" to collect literature through PubMed and Web of Science to classify disaster-related healthcare risks.

Revising Triage Rule for Post-Disaster Situation

The PHC has the following devices: weight scales, tape measures, blood pressure meters, glucose meters, body thermometers, pulse oximeters, urine test strips, and hemoglobin meters. Through examinations with these devices, a health assessment can be made. The health assessment logic, called Bangladesh logic (B-Logic), was introduced into all the disease management activities of PHC in Bangladesh [5].

In B-logic criteria, the results are divided into four stages (green, yellow, orange, and red), and they form a health assessment of Bangladeshis under non-disaster conditions. B-logic criteria do not address the post-disaster conditions because there is often a shortage of healthcare workers, and the unavailability of medicine is more serious in post-disaster than in non-disaster conditions. Therefore, we designed a triage protocol using B-logic and conducted a series of medical questionnaires on possible symptoms in post-disaster areas. The symptoms are assigned with risk assessment.

Operation Test

To examine the feasibility of the designed triage protocol, an operation test was conducted in chosen areas of Hiroshima City that were affected by a large debris flow in August 2014. This debris flow resulted in many casualties with 74 deaths and 44 injured [6]. Some victims were still living in disaster

shelters in October 2014. We performed an operation test in the post-disaster shelters of Hiroshima on October 25th to better understand their health conditions and to classify the operation, suitability, and efficiency of PHC performance in post-disaster areas.

Results

Logistics Classification Research Result

The investigation shows that in Bangladesh, floods and cyclones cause destructive damage during and after the disasters. They are the two disasters in the top 10 disasters that affected most people from 1985 to 2014 [7].

As the most frequent disaster in Bangladesh, flood disasters cause deaths from drowning (67.6%), physical trauma (11.7%), heart attack (5.7%), fire (3.6%), electrocution (2.8%), and other (8.6%) [8]. In addition to the direct impact of the disaster, health conditions of post-disaster populations are also affected by poor sanitary conditions of shelters, increasing tobacco abuse, and more stress.

Fever, diarrhea, respiratory problems, and abdominal pain are the most common symptoms in post-disaster Bangladesh. Approximately 20–40% of these subjects had diarrhea that mainly resulted from cholera and rotavirus infections [9-11]. For acute respiratory infections, a moderate increase in risk during the six months after the flood and the subsequent 18 months were found [12]. Chronic stress after natural disasters may also significantly affect cardiovascular risk factors [13]. People with NCDs are more vulnerable in emergencies and disasters, when emergencies exacerbate NCDs, leading to acute complications [14]. Natural disasters cause a higher prevalence rate of NCDs and negatively impact people with pre-existing conditions [15, 16].

Triage protocol using a new logic

The PHC algorithm was redesigned to target disaster situations. In the Disaster Logic (D-Logic) algorithm, the criteria for teleconsultation were adjusted for fewer people to see a doctor. This adjustment was acceptable in the short term, just after a disaster, to avoid congestion. A new assessment was added into D-logic for better coverage in disaster-related physical and mental health conditions.

D-logic was introduced to provide initial assessments of age and symptoms. Assessments based on sensor data and symptom questionnaire were created for a general examination of healthcare risks in our logistics classification of disasters. With the obtained results of D-logic, the remote doctor will make decision and instructions by Skype and e-prescription.

The D-logic triage was designed in the following steps:

1. Make an initial assessment with three questions shown in Figure 1 for the disaster patients.
2. According to the results of the initial assessment, disaster patients will be divided into five groups with different examinations (Figure 1).
3. Following check items and criteria are designed as D-logic_1 and 2, which are two examinations. D-logic_1 is based on the sensor measurements (Table 1). The red items in Table 1 are adjusted based on B-logic to fit with healthcare worker insufficiency in a post-disaster area. Referring to the index of symptoms questionnaire (Table 2), D-logic_2 (Table 3) will also perform a questionnaire examination with the criteria to decide if the subject needs a teleconsultation with a remote doctor. Symptoms in Table 2 will be checked based on the result of an initial assessment (Figure 1). From the data presented in Table 2, scores for each observation in Table 3 can be added to show if there is a minor or major abnormality. As a result, subjects with more than two scores will be required to have a teleconsultation with remote doctors.

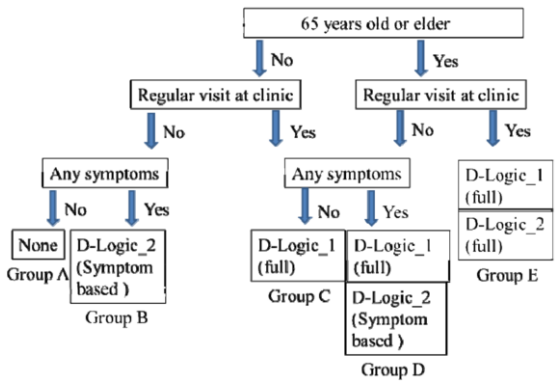


Figure 1 – Initial assessment and Flowchart of D-logic

Operation Test

To verify the operation, suitability, and turn-around time of the adapted PHC system to a post-disaster area, we tested the D-logic at the debris flow disaster shelters in Hiroshima, Japan.

We prepared the questionnaires and medical sensors to conduct the examination of two criteria, D-logic_1 and D-logic_2, to the victims who were still living in disaster shelters. Three disaster shelters and one area affected by debris flow in the city were chosen for the operation test.

We used three disaster victims as test subjects and used the medical sensors and questionnaires based on D-logic_1 and D-logic_2. Some problems and improvements to the procedure were clarified through the operation test as mentioned below.

All the subjects in our limited operation test answered that they did not feel any health problems during the initial assessment. However, through the questionnaires and examinations we found that two subjects showed some abnormalities on the examination results. Though neither of them was aware of abnormalities in their health conditions before we made the examination, one of them had a blood pressure of 158/94 mmHg, whereas the other one had a blood sugar of 150 md/dl two hours after eating. The ages of these two subjects were over 65, and this point indicates a higher risk of NCDs.

Table 1 – D-logic_1 (Red parts are the adjusted items based on B-logic, ordinal criteria)

Test	Normal	Caution	Remote medicine	Remote medicine & Encouragement to visit clinic
Body Weight Change (x kg)	x < ±1.0	±(1.0 ≤ x < 3.0) or unknown	±(3.0 ≤ x < 5.0)	±5.0 ≤ x
Blood Pressure (x mmHg systolic BP and y mmHg diastolic BP)	x < 140 (x < 130)	140 ≤ x < 160 (130 ≤ x < 140)	160 ≤ x < 200 (140 ≤ x < 180)	200 ≤ x (180 ≤ x)
	y < 90 (y < 85)	90 ≤ y < 100 (85 ≤ y < 90)	100 ≤ y < 120 (90 ≤ y < 110)	120 ≤ y (110 ≤ y)
Fasting Blood Sugar (x mg/dl)	x < 100	100 ≤ x < 126	126 ≤ x ≤ 200	200 ≤ x
Postprandial Blood Sugar (x mg/dl)	x < 140	140 ≤ x < 200	200 ≤ x ≤ 300	300 ≤ x
Urine test				
Urine Protein	- /± (-)	+ (±)	≥2+ (≥+)	
Urine Sugar	- /± (-)	≥+ (±)	(≥+)	
Urobilinogen	±		+≤ (≥+)	
Pulse Ratio	60 ≤ x < 100	50 ≤ x < 60 or 100 ≤ x < 120	X < 50 or 120 ≤ x	
Arrhythmia	None		Yes	
Number of Fresh Skin lesion	None		None	2 or more
Temperature(Celsius degree)	x < 37	37 ≤ x < 37.5	37.5 ≤ x < 38.5 (37.5 ≤ x)	38.5 ≤ x
SpO2(x %)	x ≥ 96	93 ≤ x < 96	90 ≤ x <93	x < 90
Hemoglobin (x g/dl)	x ≥ 12	10 ≤ x < 12	8 ≤ x <10	x < 8

Table 2 – Examination Item Index for D-logic_2 Check sheet

Symptom		Checksheet Number											
S1	Loss of consciousness	1	31	109									
S2	Depression	2	3	6	7	102							
S3	Flash back	3											
S4	Fever without any symptoms	101											go to S24
S5	Fever with respiratory symptoms	101											go to S11
S6	Fever with gastrointestinal symptoms	101											go to S11
S7	Fever with trauma	101											go to S21,S24
S8	Fever with other symptoms	33	101										
S9	Fatigue	1	2	5	6	7	101	102	103				go to S24
S10	Dizziness	1	4	101	103	105	106	108					
S11	Respiratory symptoms	12	13	14	15	16	17	101	105	107			
S12	Headache	1	4	8	9	31	101	103					
S13	Palpitation	6	17	19	101	105	106	108					go to S24
S14	Chest pain	19	20	103	105	106	107						
S15	Paralysis	1	4	21	22	23	31	103	105	106	107	109	
S16	Nausea or vomiting	7	8	9	10	11	25	31	101				go to S24
S17	Diarrhea	7	8	9	10	11	25	101					go to S24
S18	Stomachache	7	8	9	10	11	25	101					
S19	Abdominal pain	7	8	9	10	11	25	101					
S20	Constipation	11	26										go to S24
S21	Trauma	1	27	28	29	30	31	101					
S22	Other symptoms	33											
S23	Risk of pulmonary thrombosis	32											
S24	Dehydration	5	24	25	32	101	102	103	105				

Table 3 – D-logic_2 (Symptoms based check sheet)

D-logic2 Number	Observation	Normal (score 0)	Minor Abnormality (score 1)	Major Abnormality (score 2)
1	Confusion or loss of consciousness	No	Yes	
2	Sleep disturbance	No	Yes	
3	Flash back	No		Yes
4	Dizziness	No	Yes	
5	Thirsty	No	Yes	
6	General fatigue	No	Yes	
7	Loss of appetite	No	Yes	
8	Nausea	No	Yes	
9	Vomiting	No	Yes, only one time without bloody emesis	Yes, two times or more, or bloody emesis
10	Stomach ache	No	Yes, mild	Yes, severe
11	Abdominal pain	No	Yes, mild	Yes, severe
12	Cough	No	Yes, mild	Yes, severe
13	Nasal congestion	No	Yes	
14	Sore throat	No	Yes, mild	Yes, severe
15	Sputum	No	Yes	
16	Hemoptysis	No		Yes
17	Respiratory distress	No	Yes	
18	Headache	No	Yes, mild	Yes, severe
19	Palpitation	No	Yes	
20	Chest pain	No	Yes, mild	Yes, severe
21	Paralysis	No		Yes
22	Dysarthria	No		Yes
23	Visual disturbance	No		Yes
24	Urination	Normal	Less than normal	Anuria or oliguria (≤2/day) or much more than usual
25	Diarrhea	No	Yes, not watery	Yes, watery
26	Defecation	Normal	Less than normal	Severe (no defecation for more than 5days)
27	Surface injury	No	Yes, mild, and bleeding (-), and pus formation (-) now	Yes, severe, or bleeding or pus formation
28	Swelling	No	Yes, mild	Yes, severe
29	Joint pain	No	Yes, mild	Yes, severe
30	Pain of injury	No	Yes, tolerable without pain reliever	Yes, intolerable without pain reliever
31	Head trauma	No		Yes
32	Staying or sleeping in narrow place	No	Yes	
33	Any other symptoms	No		Yes
101	Fever (Celsius degree)	$x < 37$	$37.0 \leq x < 37.5$	$37.5 \leq x < 38.5$
102	Loss of body weight (kg)	$x < \pm 1.0$	$\pm 1.0 \leq x < 3.0$ or change is unknown	$x \geq \pm 3.0$
103	Systolic blood pressure (mmHg)	$100 \leq x < 140$	$140 \leq x < 160$ or $80 \leq x < 100$	$x < 80$ or $x \geq 160$
105	Pulse rate	$60 \leq x < 100$	$50 \leq x < 60$ or $100 \leq x < 120$	$x < 50$ or $x \geq 120$
106	Arrhythmia	No		Yes
107	Low oxygen in arterial blood (%)	$x \geq 96$	$93 \leq x < 96$	$x < 93$
108	Anemia (x g/dl by Hemoglobin meter)	$x \geq 12$	$10 \leq x < 12$	$x < 10$
109	Japanese Coma Scale	$x = 0$	$x = 1$ or 2 or 3	$x \geq 10$

During the operation test interview, some items were difficult to answer. These items are clearer to the people around the subjects. In the future, the items of the questionnaire can be divided into subjective and objective symptoms. Only the items on subjective symptoms will be questioned, and the examiner or the subjects' family can judge the items on objective symptoms. Other items such as arrhythmia or number of skin lesions were chosen as D-logic_1 and D-logic_2 components, increasing the examination turnaround time.

We conducted a full examination of D-logic and measured the turnaround time. In total, approximately 15 min on average was taken from the interview and 6 min on average was taken from the examination by medical examiners. In this operation test, the disaster ended a few months ago, which may be the reason why many victims were not fully cooperative at first. Extra time was needed to communicate with the victims for them to join the test. Months after the disaster ends, health-seeking behavior is poor. However, health-seeking behavior is prevalent earlier in post-disaster areas when possible trauma and shelter conditions are considered [17].

Among the three subjects, two had trauma due to tumbling on the second day after the debris flow. The debris flow sand deposit increased the chance of slipping. Secondary injuries in a post-disaster area should be considered when responding to the healthcare risks after disasters that cause wet and slippery conditions.

Discussion

Limitation of the System

Disaster damages may lead to a power outage and network malfunction, and this will affect the PHC consumables.

Battery backup may be too limited to cover all patients in a post-disaster area. To relieve the situation, an initial assessment can be used to screen the victims with the most health risks and relieve the pressure on consumables, resources, and healthcare worker insufficiency. However, an operation test on a large number of subjects is necessary in the future to improve the efficiency of triage by D-logic.

In Bangladesh, pharmacies and drugstores are common even in rural areas. Para-professionals, pharmacy, and drugstore sales people were the major healthcare providers to disaster-affected people after Cyclone Sidr [18]. Necessary medicine can be prepared in advance from the drugstores in Bangladesh and help the patients in the worst condition. However, treatment for PHC in Japan is limited to emergency usage. Different preparations should be planned depending on the policies of different countries and situations.

Further Improvement

Disasters have impacts on critical infrastructure leading to power outrage and network failure. PHC devices are vulnerable to post-disaster power outage and wet conditions. Backup for device consumables and waterproof measurement are necessary for future use of the post-disaster condition. During the operation test in Hiroshima, questionnaires data and examination were manually recorded. In the future, PHC will automatically switch from B-logic to D-logic by information technology, and this will result in less turnaround time and more efficiency in post-disaster settings.

Conclusion

PHC is a disease prevention program in Bangladesh and has undertaken health checkups for more than two years. It is being developed to reach the low-income people. Its mobility and agility make it easier to be carried into the rural and post-disaster areas. With widespread use in ordinary conditions, the emergency mode can become more suitable and rapidly introduced in post-disaster areas. Existing packages in disaster areas can be switched into emergency mode immediately after the disaster ends to respond rapidly to the post-disaster health risks.

References

[1] World Health Organization (WHO): The World Health Report 2006 - Working together for health. Geneva: World Health Organization; 2006.

[2] Cash RA, Halder SR, Mushtuq H, et al. Reducing the health effect of natural hazards in Bangladesh. The Lancet. 2013; 382: 2094-2103.

[3] Ochi S, Nakagawa A, Lewis J, Hodgson S, Murray V. The great East Japan earthquake disaster: distribution of hospital damage in Miyagi Prefecture. Prehosp Disaster Med. 2014; 29(3): 245-53.

[4] Ohkouchi S, Shibuya R, et al. Deterioration in regional health status after the acute phase of a great disaster: respiratory physicians' experiences of the Great East Japan Earthquake. Respir Investig. 2013; 51(2): 50-5.

[5] Nohara Y, Kai E, Ghosh P et.al. A Health Checkup and Tele-Medical Intervention Program for Preventive Medicine in Developing Countries: A Verification Study. Journal of Medical Internet Research, 2014, in press.

[6] Damage assessment report on the heavy rain of August 19th. Disaster Prevention Center of Hiroshima. www.pref.hiroshima.lg.jp/uploaded/attachment/139124.pdf Accessed November 21th, 2014.

[7] Country Report: People's Republic of Bangladesh. Asian Disaster Reduction Center Web site. http://www.adrc.asia/countryreport/BGD/2013/BGD_CR2013B.pdf. Accessed October 7th, 2014.

[8] Jonkman SN, Kelman I. An analysis of the causes and circumstances of flood disaster deaths, Disasters 2005; 29: 75-97.

[9] Kunii O, Nakamura S, Abdur R, Wakai S. The impact on health and risk factors of the diarrhea epidemics in the 1998 Bangladesh floods, Public Health. 2002; 116: 68-74.

[10] Siddique AK, Baqui AH, Eusof A, Zaman K. 1988 Floods in Bangladesh, Pattern of Illness and causes of Death, J Diarrhoeal Dis Res. 1991; 9(4): 310-314.

[11] Schwartz BS, Jason B, Diarrheal epidemics in Dhaka, Bangladesh, during three consecutive floods: 1988, 1998, and 2004, Am. J. Trop. Med. Hyg. 2006; 74(6): 1067-1073.

[12] Milojevic A, Armstrong B, Hashizume M, et al. Health Effects of Flooding in Rural Bangladesh, Epidemiology. 2012; 23: 107-115.

[13] Jiao Z, Kakoulides SV, Moscona J, et al. Effect of Hurricane Katrina on Incidence of Acute Myocardial Infarction in New Orleans Three Years After the Storm, Am J Cardiol. 2012; 109: 502–505.

[14] Demaio A, Jamieson J, Horn R, Courten MD, Tellier S. Non-Communicable Diseases in Emergencies: A Call to Action, PLoS Currents Disaster. 2013; 5: 1-8.

[15] Konno S, Hozawa A, Munakata M. Blood pressure among public employees after the Great East Japan Earthquake: the Watari study. Am J Hypertens. 2013; 26(9):1059-63.

[16]Tomio J, Sato H, Mizumura H. Interruption of medication among outpatients with chronic conditions after a flood. Prehosp Disaster Med. 2010; 25(1):42-50.

[17] Daniels A, Chapin E, Aspilcueta D, Doocy S. Access to health services and care-seeking behaviors after the 2007 Ica earthquake in Peru. Disaster Med Public Health Prep. 2009; 3(2): 97-103.

[18]Jalal U, Robert EM. Socioeconomic factors differentiating healthcare utilization of cyclone survivors in rural Bangladesh: a case study of cyclone Sidr. Health Policy Plan. 2014; 29 (7).

Address for correspondence

humin@med.kyushu-u.ac.jp 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, JAPAN

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-84

Assessing the Potential Use of Eye-Tracking Triangulation for Evaluating the Usability of an Online Diabetes Exercise System

Clara Schaarup[a], Gunnar Hartvigsen[b,c] , Lars Bo Larsen[d], Zheng-Hua Tan[d], Eirik Årsand[c], Ole Kristian Hejlesen[a,b]

[a] *Department of Health Science and Technology, Aalborg University, Denmark*
[b] *Department of Computer Science, University of Tromsø – The Arctic University of Norway, Norway*
[c] *Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø*
[d] *Department of Electronic Systems, Aalborg University, Denmark*

Abstract

The Online Diabetes Exercise System was developed to motivate people with Type 2 diabetes to do a 25 minutes low-volume high-intensity interval training program. In a previous multi-method evaluation of the system, several usability issues were identified and corrected. Despite the thorough testing, it was unclear whether all usability problems had been identified using the multi-method evaluation. Our hypothesis was that adding the eye-tracking triangulation to the multi-method evaluation would increase the accuracy and completeness when testing the usability of the system. The study design was an Eye-tracking Triangulation; conventional eye-tracking with predefined tasks followed by The Post-Experience Eye-Tracked Protocol (PEEP). Six Areas of Interests were the basis for the PEEP-session. The eye-tracking triangulation gave objective and subjective results, which are believed to be highly relevant for designing, implementing, evaluating and optimizing systems in the field of health informatics. Future work should include testing the method on a larger and more representative group of users and apply the method on different system types.

Keywords:

Eye Tracking; Usability; Diabetes Mellitus; Post-Experienced Eye-Tracked Protocol; Consumer Health Information.

Introduction

Consumer health information technology (CHIT) is a new term in the healthcare sector, which refers to a computer-based system that facilitates access to information and behaviour changes that promote health and well-being [1]. Recently, a systematic review by Jenni Cornelly et al. examine the effectiveness of technology to promote physical activity in people with Type 2 diabetes [2]. Their findings indicate that technology-based interventions to promote physical activity are effective. Nevertheless, they demonstrate a need for evidence of the sustainability of the technology [2]. In a systematic review by Yu et al. they also identify the needs of sustainability, effectiveness, usefulness, and usability regarding CHIT and management of diabetes [3].

An example of a CHIT-system, which has been tested for usefulness and usability, is the web-browser based patient health IT system named "The Online Diabetes Exercise System". The system was developed to motivate people with Type 2 diabetes to do a 25 minutes low-volume high-intensity interval training program [4]. The development of this CHIT-system was motivated by the fact that in 2014, 387 million people worldwide had diabetes and that the number is expected to rise to more than 592 million by 2035 [5].

In a previous multi-method evaluation of the CHIT-system, several usability issues were identified and corrected [4]. Despite the thorough testing, it is unclear whether all usability problems were identified using the multi-method evaluation.

To assess the amount of usability problems that were not identified in the multi-method evaluation, we have applied an extended version of eye-tracking, named Eye-tracking Triangulation, to the CHIT-system. Eye-tracking triangulation is made of two successive sessions. The first session of conventional eye-tracking with predefined tasks followed by the second session: The Post-Experience Eye-Tracked Protocol (PEEP) [6]. Eye-tracking is a technique where eye movements are recorded while users look at a stimulus [7]. In PEEP, the users have to explain their decisions and thoughts while a retrospective replay of their eye-tracking data is shown [6]. Even though eye-tracking has been used in psychology for decades, to our knowledge, eye-tracking triangulation has not been applied to health informatics systems.

Our hypothesis is that adding the eye-tracking triangulation to the multi-method evaluation increases the accuracy and completeness when testing the usability of CHIT-systems. The objective of the present study was to assess the potential use of eye-tracking triangulation for evaluating the usability of a CHIT-system, The Online Diabetes Exercise System, which, before the present evaluation, had been optimized using multi-method evaluation.

Materials and Methods

The Patient Health IT System

As already outlined, the eye-tracking triangulation was applied to the Online Diabetes Exercise System (Figure 1). The system is a web-browser based prototype and was designed, implemented, evaluated and optimized using the iterative, systematic, and holistic multi-method evaluation, which consisted of interviews, paper prototyping, heuristic evaluation and tests with users [4]. The system contains two major functions: a 25 minute low-volume high-intensity interval training

program (HIT program), which the users in this study only looked at, but did not perform, and a glucose diary.

Figure 1 - The Menu of the Online Diabetes Exercise System

Recruitment

Eight users were recruited for the study - five women and three men. The background of the participants were diverse, ranging from having short to long educations: a biomedical laboratory technician with a master in health informatics, an occupational therapist, a global IT-manager, a nurse, a boilermaker, a shipper, an MA in Danish and Psychology, and a veterinary nurse. Their age ranged from 34 to 69 years. Six out of the eight had a Body Mass Index (BMI) of more than 25 and were per definition classified as being overweight. All the users exercised daily, they were native Danish speaking people, and they reported to use the Internet on a daily basis. Five out of the eight users wear glasses, which is not a problem while tracking their eye movements.

Equipment

The users' eye movements were tracked using a Tobii Eye-Tracker X120. The eye-tracker was placed below a Samsung Monitor 21.5 inch monitor used by the test participants, as shown in Figure 2. All eye movements were recorded using the Tobii Studio software and stored together with a screen video recording, voice input, and mouse clicks for subsequent analysis [8]. Eye-movements consist of fixations and saccades. A fixation is defined as a moment where the eyes are almost motionless and it generally has a duration of 100 ms to 500 ms. A saccade is defined as a quick movement between different fixations and on average it lasts for about 250 ms, for example while reading [6]. In this study, we were only interested in fixations periods. While the test participants sat in front of the eye-tracker, their eye movements were shown simultaneously on a separate monitor, allowing the facilitator to follow the experiment and ensuring the equipment was capturing the data at all times (to the left in Figure 2). Prior to every test session, the equipment was calibrated to each participant to ensure optimal accuracy (about 0.5 cm on a monitor placed 70 cm in front of the test participant).

Study Design and Procedure

Session 1: Eye-Tracking

Initially, each user received an introduction to the purpose of the study. They then received instructions on how to seat themselves in a comfortable position in front of the Tobii Monitor, while completing the following tasks:

1. Set up an account in the Online Diabetes Exercise System
2. Search for help due to forgotten login information
3. Login to the Online Diabetes Exercise System
4. Acquire knowledge about the Online Diabetes Exercise System through an available user guide
5. Start and finish the low volume HIT program (not performing the actual training)
6. Register and review information in the glucose diary
7. Log out from the Online Diabetes Exercise System

After assenting to the instructions and the tasks, the facilitator calibrated the Tobii eye-tracker and the users began completing the tasks.

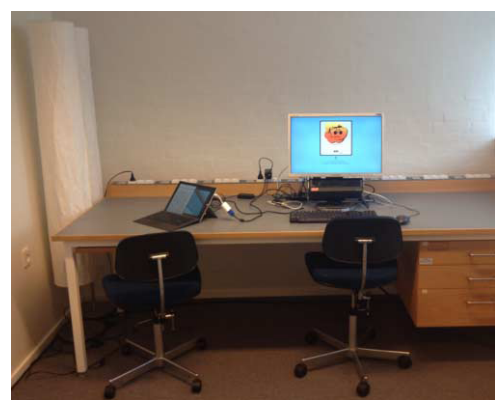


Figure 2 - The eye-tracking set up

Session 2: Post-Experience Eye-Tracking Protocol (PEEP)

A dicta-phone was turned on to collect the users' answers. The facilitator had predefined six Areas of Interests (AOIs) which would be the basis for the PEEP-session. The AOIs were: (1) *Information Text;* (2) *Forgot login/Create an Account;* (3) *Login;* (4) *Subheadings;* (5) *Colour Box;* (6) *Progress Bar.* The AOIs were spread over several different screens. The users were asked to provide retrospective protocols, meaning that while they were shown a replay of their eye-tracking data they were asked to explain their thoughts and decisions to each AOI [6]. The following details were specifically queried: long fixations; text scanning rather than reading, and failing to look at specific elements that were considered important for the task. At the end of the session, the users were asked if they had any further comments, and if not, Session 2 was complete. Session 1 took approximately 10 minutes and Session 2 took 15-20 minutes.

Data Analysis

Statistical data about the six AOIs were extracted from the eye tracking software regarding the two metrics shown in Table 1. Usability issues were identified from the statistical data. Valuable and complementary inputs to analyse the statistical data were provided by the PEEP-technique. The PEEP inputs were afterwards compared with the identified usability issues.

Table 1 – Eye-Tracking Metrics

Metric	Explanation
Time to First Fixation	The time in seconds from when the stimulus was shown until the start of the first fixation within an AOI.
Visit Count	Number of times the user fixate on an AOI.

Results

The five heatmaps shown in Figures 3 to 7 illustrate the visual behaviour of users during the eye-tracking sessions. The colour coding illustrates the total time spent looking at various areas. Green equals a shorter time, yellow medium time, and red longer time. Dashed rectangles indicate AOIs.

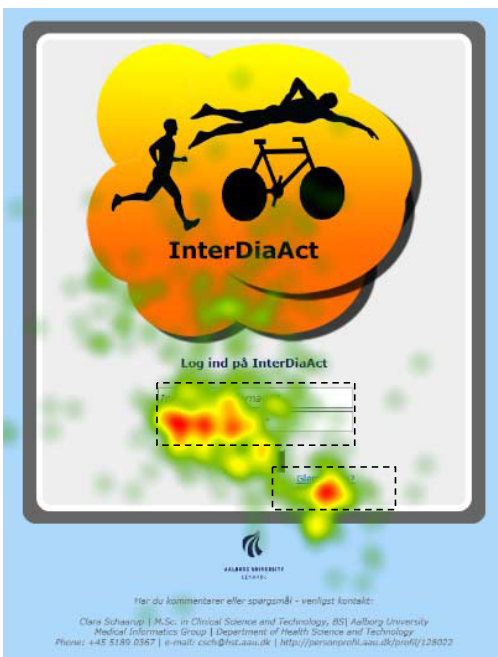


Figure 3 – The welcome screen has two AOIs: the 'Login' area and the area for 'Forgot login/Create an Account'

Figure 3 shows how the users fixated at the AOI-*Login* and the AOI-*Forgot login/Create an Account*. On average for the Login AOI, the users had 12.00 visit counts and spent 1.38 seconds before their first fixation, which can be seen in Table 2. During the PEEP session, some of the users reported that the font used was too small to identify the button 'Create an account'. Later in the session, they said that they just did not look carefully enough.

Table 2 – The six predefined areas of interest correspond to the two metrics: Visit Count & Time to First Fixation

Area of Interest; (AOI)	Visit Count; Mean(count)	Time to First Fixation; Mean(.seconds)
Information Text	10.25	17.80
Forgot login/Create an Account	5.17	10.18
Login	12.00	1.38
Subheadings	12.75	2.96
Colour Box	7.20	0.25
Progress Bar	1.00	1.33

The heatmap in Figure 4 illustrates how the users read the information about the newest research regarding low-volume high-intensity interval training. As can be seen in Table 2, the users on average had 10.25 visit counts on this AOI, the *'Information Text'*, and spent 17.80 seconds before they had their first fixation on it. In the PEEP session, the users reported that it was overwhelming with all the information on this AOI and was hard to read it all. In addition, they reported that the font size was too small. Nevertheless, some of the users mentioned that it was interesting to read. A later test of the text showed a readability index of 49, which corresponds to very difficult readings such as academic papers [9].

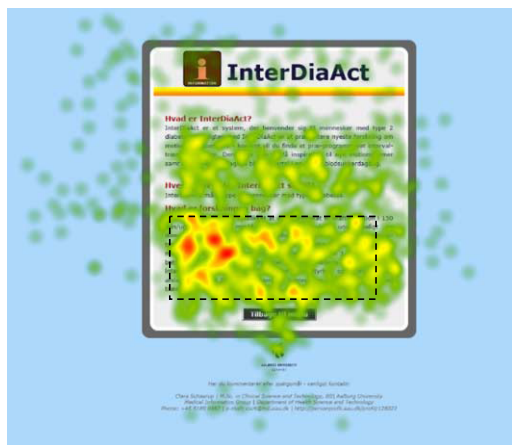


Figure 4 –AOI with the newest research regarding low-volume high-intensity interval training

Figure 5 shows the heatmap for the *Progress Bar*. On average, the users had 1 visit count and spent 1.33 seconds to identify the AOI. In the PEEP session, the users reported that the progress bar was too small and unclear.

Figure 5 – While the low-volume high-intensity interval training program was running, the users were presented to a red progress-bar placed in the top of the of the screen

Figure 6 shows the heatmap for the page where users can obtain information on how to add another training program or

switch to a different type of exercise, for example from biking to swimming. The page is divided in subheadings and one of these is defined as an AOI. On average, the users had 12.75 visit counts and 2.96 seconds to first fixation on the AOI *Subheadings*. During the PEEP session, the users reported that the subheadings helped them to get an overview of the information.

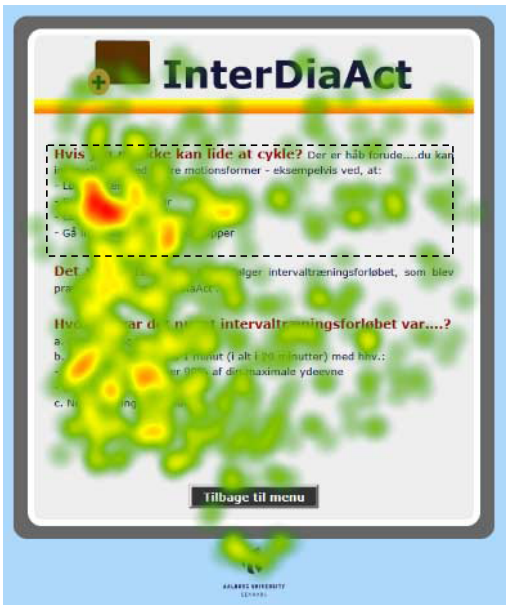


Figure 6 - The heatmap of the page where users can get information on how to add another training program or switch to a different type of exercise

Figure 7 shows the information explaining the colour scheme used during the training program – green for initial warm-up and final cool-down period, yellow for pauses between bursts of high intensity training, and red for the short periods with high intensity activity. The illustration of this at the bottom of the figure has been chosen as an AOI, the *Colour Box*. On average, the users had 7.2 visit counts and spent 0.25 seconds to first fixation at the AOI. In the PEEP session, the users reported that they preferred figures rather than text.

Figure 7 – The heatmap of the page explaining the colour scheme used during the training program

Discussion

The aim of the study was to assess the potential use of eye-tracking triangulation to evaluate the usability of the Online Diabetes Exercise System. The eye-tracking triangulation was made up of two successive sessions: The first session of conventional eye-tracking with predefined tasks followed by the second session, The Post-Experience Eye-Tracked Protocol (PEEP), the latter being based on the eye-tracking heatmaps and the six selected Areas of Interest [6].

Eye-tracking triangulation is a technique, which has been used for more than 20 years in usability studies for marketing purposes, and which is well known in cognitive psychology [7]. The technique is an attractive and powerful tool because it delivers objective results such as heatmaps and subjective results such as statements and reports that clarify potential misinterpretations of the heatmaps. The combination of objective and subjective results can be argued to make the method more valid and reliable.

In our study we saw the strength of the eye-tracking triangulation on several occasions, where the method detected usability issues that had not been found with the previously applied multi-method evaluation consisting of interviews, paper prototyping, heuristic evaluation and tests with users. For instance, we had an assumption that the participants would use the progress bar while they were looking at the low-volume high-intensity interval training program (Figure 5). However, the results showed that the users on average only had one visit count at this AOI and thus basically did not use it (Table 2). In the following PEEP session, we received an expanded explanation for this. This is an example of how eye-tracking triangulation, in our study, provided relevant information that was not provided by any of the four tests and evaluations in the multi-method evaluation.

Another example, which illustrates the strength of the triangulation, is the heatmap in Figure 4, with an AOI describing the newest research regarding low-volume high-intensity interval training. From the heatmap, we saw that the users read the information text, i.e., looked at the AOI labelled *Information Text*, but also that they looked at areas far away from the AOI several times. The heatmap did not tell us why, but the users explained this during the following PEEP session and thereby gave important information on a very relevant problem – that it was overwhelming with all the information on this AOI and that it was hard to read all. Again, this problem had not been detected by any of the four tests and evaluations in the multi-method evaluation.

The two examples illustrate how the combination of the eye-tracking and the post-experienced eye-tracked protocol can increase the accuracy and completeness of findings of usability issues in a consumer health information technology system.

From the literature, we know that it may be difficult to evaluate systems in the field of health informatics [10]. Bürkle et al. have in their study tried to assess advantages and disadvantages of different study designs, which may be potential methods to evaluate health informatics. They looked at the following designs: a randomized controlled study, a controlled study, a non-controlled study, and a simulation study and they found that evaluating health informatics depends on: the goal of the evaluation, available resources, human factors, and what type of technology that has to be examined. They concluded that a mixed approach that combines measurement of several indicators is the optimal way to evaluate health infor-

matics [10]. This conclusion is supported by the findings in the present study, which, in addition, indicates a potential added value of including eye-tracking triangulation in a mixed approach to evaluate health informatics.

Despite the promising results in our study, there were some limitations in the design. Firstly, we only used eight participants and they were probably not representative of all potential users of such a system. Secondly, the eye-tracking triangulation was only applied to one system, the Online Diabetes Exercise System, and the results may be different when applying the method to other types of consumer health information technology.

Conclusion

In conclusion, several problems, not identified previously, were found using eye-tracking – adding eye-tracking triangulation to the multi-method evaluation was found to increase the accuracy and completeness. Eye-tracking triangulation gave objective and subjective results, which are believed to be very relevant for designing, implementing, evaluating and optimizing systems in the field of health informatics.

Even though eye-tracking has been used in other areas for decades, to our knowledge, eye-tracking triangulation has not been applied to health informatics systems. Our findings suggest a future role for eye-tracking triangulation when evaluating systems in the field of health informatics.

Future work should include more studies of the potential benefits from using eye-tracking. The method should be assessed using larger and more representative groups of users and the method should be applied to a variety of different types of systems.

Acknowledgements

The authors would like to thank the users who participated in the eye-tracking triangulation sessions. The work was supported by the 7th framework EU project eWall.

References

[1] Lai YH. A Meta-Analysis on the Information Technology Application in Patients' Knowledge and Anxiety in Taiwan. Asian Journal of Applied Science. 2014;02(02):138-144.

[2] Conelly J, Kirk A, Masthoff J, MacRury S. Systematic Review or Meta-analysis. The use of technology to promote physical activity in Type 2 diabetes management a systematic review. Diabetic Medicine. 2013; 30. 1420-1432.

[3] Yu CH, Bahniwal R, Laupacis A. Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers. J Am Med Inform Assoc 2012;19:514-522.

[4] Schaarup C, Hejlesen OK. A Multi-Method Pilot Evaluation of An Online Diabetes Exercise System. MIE 2015 Proceedings, accepted, pp.5.

[5] International Diabetes Federation. IDF Diabetes Atlas Sixth Edition. International Diabetes Federation, 2013.

[6] Ball LJ, Eger N, Stevens R, Dodd J. Applying the PEEP method in usability testing. Interfaces, 67 (summer 2006), 15-19.

[7] Lai ML, Tsai MJ, Yang FY, Hsu CH, Liu TC, Lee SWY, Lee MH, Chiou GL, Liang JC, Tsai CC. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educational Research Review, Elsevier, 2013:10: 90-115.

[8] Tobii. User Manual – Tobii Studio, Version 3.2. Manual Version 3.2 Rev A. 11/2012. Tobii Technology AB.

[9] Anderson J. Lix and Rix: Variations on a Little-known Readability Index. Journal of Reading, 1983;26:6, 490-496.

[10] Bürkle T, Ammenwerth E, Prokosch HU, Dudeck J. Evaluation of clinical information systems. What can be evaluated and what cannot? Journal of Evaluation in Clinical Practice, 2001; 7:4, 373-385.

[11] Friedman CP, Wyatt J. Evaluation Methods in Biomedical Informatics. Springer. 2006.

[12] Brender J. Handbook of Evaluation Methods for Health Informatics. Elsevier Academic Press: Burlington, MA. 2006.

[13] Guideline for good evaluation practice in health informatics (GEH-HI). https://www.ghdonline.org/uploads/Eval_2011_NykanenEtAl_GuidelineGoodEvalPracHlthInfo_GEP-HI_IMJI.pdf

Address for correspondence

Clara Schaarup, Medical Informatics Group, Department of Health Science and Technology, Aalborg University, Frederik Bajers Vej 7 C2-221, DK-9220 Aalborg, Denmark. E-mail: csch@hst.aau.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-89

Mobile App to Reduce Inactivity in Sedentary Overweight Women

Joseph Finkelstein, McKenzie Bedra, Xuan Li, Jeffrey Wood, Pamela Ouyang

Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

Recent studies demonstrated that the duration of inactivity (sedentary state) is independently associated with increased risk of cardiovascular disease. Our goal was to develop the technology that can measure the amount of inactivity in real time, remind a person that a preprogrammed period of inactivity has occurred and encourage a period of activity, and provide web-based feedback with tailored information to the participant and investigators. Once it was developed, we carried out a pilot study in a group of sedentary overweight women. The objective of the study was to assess potential of the mobile app to reduce inactivity in our target population. A randomized crossover design was employed with study subjects randomly assigned to a 4-week each "message-on" and "message-off" periods. Out of 30 enrolled subjects, 27 completed the study. The average age of particpants was 52±12; BMI: 37±6; 47% were white and 47% were African American. Overall, inactivity was significantly lower ($p<0.02$) during "message-on" periods (24.6%) as compared to the "message-off" periods (30.4%). We conluded that mobile app monitoring inactivity and providing a real-time notification when inactivity period exceeds healthy limits was able to significantly reduce inactivity periods in overweight sedentary women.

Keywords:

Inactivity, mobile app, eHealth, telemonitoring

Introduction

Cardiovascular disease (CVD) is the leading cause of death in women in the USA. The risk factors contributing to CVD include obesity (strongly associated with insulin resistance), diabetes, inflammation, and dyslipidemia. Prevalence of CVD is lower in women than in men; however, women with diabetes carry a higher risk for CVD. Studies suggest that women with diabetes mellitus (DM) without known CVD have a greater mortality than women without DM but with known CVD.[1] Exercise reduces obesity and visceral adiposity, increases peripheral muscle glucose utilization, and increases insulin sensitivity.

The US federal guidelines recommend a daily moderate-intensity physical activity for 30 minutes. Despite these recommendations, the prevalence of obesity and diabetes continues to increase, and it is predicted that US obesity may reach 42% by 2030. In addition to the well established link between physical activity and lower CVD risk, it is now becoming evident that the duration of inactivity (sedentary state) is also associated with increased risk. The risk associated with being sedentary is independent of the amount of physical activity; i.e., the effects of too much sitting rather than too little exercise also impact CV risk.[2] Studies have generally shown that people admitting to a higher number of hours of sitting have greater risk for developing diabetes, cardiovascular disease, and all-cause mortality.[3] Thus it is plausible that reducing the amount of sedentary time may reduce CVD and diabetes risk. Furthermore, the data suggest that even active individuals would benefit from reducing their hours of inactivity. This may be particularly important in women who have greater barriers to exercise.[4]

Our goal was to develop the technology that can measure the amount of inactivity in real time, remind a person that a preprogrammed period of inactivity has occurred encourage a period of activity, and provide a web-based feedback with information tailored to the participant and investigators. Once it was developed we carried out a pilot study in a group of sedentary overweight women. The objective of the study was to assess potential of the mobile app to reduce inactivity in our target population.

Materials and Methods

For inactivity monitoring we employed a physical activity tracking device (Fitbit) and an Android smartphone with digital data plan. Fitbit uses a three-dimensional accelerometer to monitor physical activity including number of steps. Fitbit is a miniature wearable device which communicates with a smartphone via Bluetooth connection to transmit the physical activity data. The physical activity data are then relayed by a smartphone to a Fitbit database hosted by the company, which provides developers with API to access the user data by custom applications.

System Design

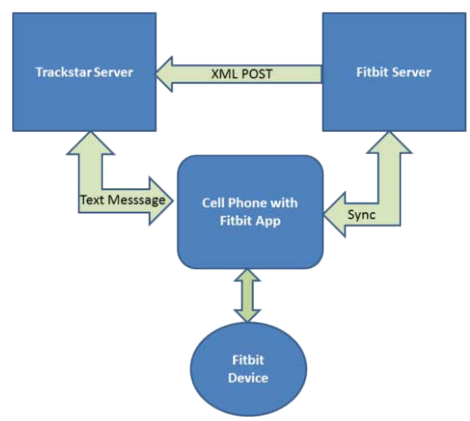


Figure 1 – Mobile App Design

The inactivity monitoring system consists of two separate programs hosted on the same server – a step monitoring app that collects data from the Fitbit users and stores it in our database, and another monitoring program that generates

tailored text messages in certain situations. Both programs were developed in C# using the Fitbit API.

The step monitoring website receives POSTed XML updates from the Fitbit server whenever the Fitbit device syncs. Each Fitbit account has a subscription service activated that tells it where to POST the XML file. The XML file is sent to our site by Fitbit automatically for each registered user account. The Fitbit device syncs every 10-15 minutes when it is in the range, which defines the maximal frequency of updates received by our server in real time. Each post provides an update on the number of steps made by a particular user in an incremental manner. The step counts starts from zero every day at 00:00am.

When the site receives the XML file it records the time and number of steps which are queried for the user using the Fitbit API. It then checks to see if there have been less than 15 steps in the past hour. If there were, and no blackout conditions apply, it will send a tailored text message to the user's phone informing that sedentary period exceeded healthy limits, and encourages the user to take a break from the sedentary position. The tailored message provides suggestions from the message library on ways to have short physical activity breaks in an office environment or at home, depending on the time of the day.

The monitoring program has four different execution modes depending on the argument(s) passed. One of those modes is the sync monitor. The sync monitoring program goes through all users and checks their last sent record, which is the last time they synced. If they have not synced for an hour or more, it should send them a text message saying so, unless such a message was sent in the past hour. This mode runs every 15 minutes. This mode addresses instances when connection between Fitbit and the cell phone is lost.

The second mode is the exercise link program, which sends a link to relevant exercise education materials from the database to every user in the database every time it is run. It should cycle through all the links in the table in order and goes back to the first one after it gets to the end. The current link is stored in the database. This program is scheduled to run at 8am every Saturday and Sunday. This mode promotes user engagement in healthy lifestyle activities by empowering the user with helpful information during weekends when users have more time to review educational messages.

The third mode is the daily report program, which sends a summary of the user's steps from the previous day. It uses the Fitbit API to query a user's total steps from the previous day. This mode provides useful feedback on a daily basis helping users monitor their activity level and adjust it when necessary.

The final mode is the no sync text. This program checks whether or not a user has synced in the last 24 hours. If not, the user is sent a text message telling them this and encouraging them to reestablish connection. This mode allows to identify potential technical issues in a timely manner.

There are blackout conditions for which no messages are sent to the patient. The first blackout condition is if the user has texted "S[X]", – where X is the number after S – so that the mobile app stops sending text messages for the next X hours. If the user texts "Okay", it does not send texts for the next 1 hour. They also do not receive texts if there are blackout times corresponding to the current time set in the database for that day. Blackout hours were entered for each patient at enrollment based on their personal preferences. This mode addresses specific user preferences in terms of their schedule when they cannot use the phone.

The design of the system has been informed by two focus groups conducted in overweight sedentary women at the beginning of the development process. The app specifications tailored participants' suggestions voiced in the focus groups.

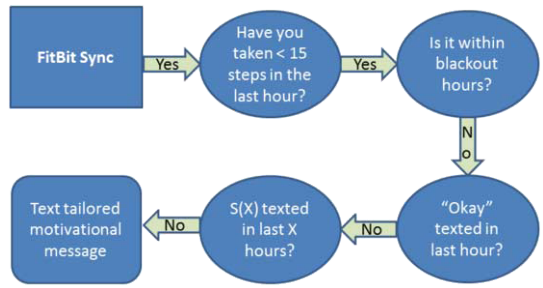


Figure 2 – Text Messaging Algorithm

Evaluation Design

A randomized crossover design was used to evaluate potential impact of the mobile app on the inactivity level (Figure 3). For this pilot we enrolled overweight sedentary women who met the following criteria.
Inclusion criteria:

- Adult women with BMI > 30 kg/m^2 who are inactive for > 3 hours on an average day

Exclusion criteria:

- Pregnant per participant's information
- Inability to walk
- Medical reasons to limit activity; e.g., unstable cardiac conditions such as angina or heart failure
- Poorly controlled hypertension, SBP >160 mm Hg, DBP >100 mg Hg (whether or not they are on antihypertensive agents)

No clinical care was provided by the research team. Participants were recruited by flyers and from previously conducted focus groups that were designed to provide feedback on the devices and messages that were used in the study. Almost all the focus group participants expressed interest in being in the study.

During the first visit the participants had their weight, height, and BP measured and blood drawn for fasting glucose and insulin. A questionnaire was administered to capture demographic and major clinical conditions to ensure that activity is possible and advisable using the American Association of Sports Medicine criteria.[5,6] Participants also completed a number of questionnaires that capture their level of activity/inactivity at work and during leisure time, as well as their readiness toward starting to increase activity. All study subjects received a Fitbit One and a smartphone. They were instructed on the use of these devices. Digital data plan and phone service were provided with the smartphone for all participants for the duration of the study. During the first 2 days (Day -2 to Day 0) participants were asked to use the devices and become comfortable with them and they were encouraged to contact the coordinator if they had any questions. The first 2 days were a regular work days and formed a baseline period of activity and inactivity. Participants were randomized to one of the two groups (Group A or B). Group A participants had the inactivity reminder system activated for time period 1 (4 weeks duration), and that function was inactivated for the second 4 weeks of the study. Group B had the inactivity reminder system inactivated for the first 4 weeks, and activated for the second 4 weeks of the

Table1 – Inactivity analysis: inactivity is expressed as fraction of the day from 8:00 am to midnight

	Inactivity of Message-off period				Inactivity of Message-on period				P-value
	Mean	SD	Min	Max	Mean	SD	Min	Max	
Total (N=27)	0.3044	0.1853	0.1200	0.9000	0.2456	0.1394	0.0700	0.6400	0.0207*
Group A (N=15)	0.3213	0.2274	0.1200	0.9000	0.2153	0.1366	0.0700	0.5200	0.0037*
Group B (N=12)	0.2833	0.1204	0.1500	0.5600	0.2833	0.1393	0.1500	0.6400	0.9824

Table2 – Analysis by steps from 00:00 am to midnight

	Steps of Message-off period				Steps of Message-on period				P-value
	Mean	SD	Min	Max	Mean	SD	Min	Max	
Total	5977.03	2108.79	2490.6	11109.9	5684.37	2148.4	2537.85	12402.61	0.5902
Group A	5614.77	1855.12	2620.55	9655	6199.11	2062.1	2969.21	9386.21	0.3303
Group B	5771.37	2552.4	2537.85	12402.6	5699.43	2224.36	2490.6	11109.89	0.8501

study. At the end of the study (Day 56), the biometric measurements and questionnaires were repeated.

In addtion to tailored text messages generated by the mobile app, the participants were allowed to see all the measurements of activity that are routinely captured and displayed by the commercial Fitbit website. This includes the number of steps per day, the number of steps climbed, distance walked, and calories burned.

At the end of the study, participants were asked their views on the devices, messages, and the effect of these on their perceptions or knowledge about activity and inactivity, the impact of these technologies on their readiness to be active, and their thoughts on whether they would be willing to continue using these devices/software for longer durations. This information will be used to improve mobile app design for a larger study.

The primary outcome of interest was the number of episodes of prolonged inactivity (> 2 hours duration) per day during the time period that the inactivity reminder was active compared to not active (time Period 1 vs 2).

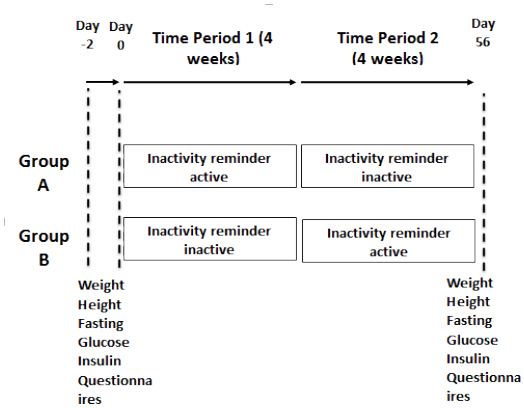


Figure 3 –Study duration and number of study visits required of research participants.

Results

Overall results were positive. Out of 30 enrolled subjects, 27 completed the study. The average age of particpants was 52±12, with average BMI of 37±6.; 47% of the participants were white and 47% were African American. There was no significant difference in baseline characteristics between Group A and B.

The inactivity was calculated as percent of consecutive 2-hour slots between 8am and midnight (total of 8 2-hr slots per day) during which the number of steps did not exceed 20 (Table 1). Overall, inactivity was significantly lower ($p<0.02$) during "message-on" periods (24.6%) as compared to the "message-off" periods (30.4%). For Group A, the mean period of inactivity as a percentage was 21.5% during the "message-on" period, and during the "message-off" period inactivity increased to 32.1% ($p<0.004$), indicating a decrease in inactivity when receiving text messages from the system. Group B, which received no tailored messaging during the first 4 weeks, showed no change when it was switched to messaging in the subsequent 4 weeks, with a mean inactivity period of 28.3% for both the "message-on" and "message-off" periods.

An analysis of the step count revealed that Group A had a higher average daily number of steps during "message-on" period (6199.1±2062.1) as compared to the "message-off" period (5614.8±1855.1). A similar but less pronounced tendency was documented in Group B with mean number of daily steps going from 5771.4±2552.4 during "message-off" period to 5699.4±2224.4 during "message-on" period.

After completion of the 8-week follow-up period, two focus groups were conducted with the study participants. A majority of the participants expressed high acceptance of the mobile app and indicated willingness to use it in the future.

Discussion

A mobile app monitoring level of inactivity in real-time has been successfully introduced in this study. The impact of tailored messaging generated by the mobile app in response to prolonged periods of inactivity has been studied in overweight sedentary women using randomized crossover design. The results of the study demonstrated a significant impact of tailored messaging on inactivity. Comparison between "message-off" and "message-on" periods showed that there was statistically significant reduction in inactivity duration when the study particpants were receiving text messages indicating that sedentary periods exceeded healthy limits and encoraging the study subjects to move. Further analysis demonstrated that significant reduction in inactivity occurred only in Group A which receved tailored messaging during the

first 4 weeks of wearing Fitbit. Participants in Group B who started receiving tailored messages only after 4 weeks of wearing Fitbit did not demonstrate decrease in inactivity after they were switched to tailored messaging. This conforms to previously described phenomenon of technological imprinting when initial patterns of new technology use are maintaned regardless of the subsequent changes in this technology.[7-9] Any change in user behavior afterward requires additonal retraining.[10,11] Our results underscore the importance of introducing fully functional mobile apps including tailored messaging from the very beginning of the intervention aimed to reduce inactivity.

Recent studies have demonstrated the importance of patient-centered delivery of the medical care tailored to individual needs, preferences, and values of the patients.[12-14] In concordance with these findings, successful implementation of the mobile app described in this study can be attributed to tailoring its specifications to preferences of the target audience and employment of participatory design principles from the beginning of the mobile app's development.[15-17] Another factor affecting high acceptance of the mobile app is that smartphones have become a ubiquitous appliance widely used in our target population.[18,19] High utility of mobile phones for health empowerment and engagement has also been demonstrated previously.[20,21]

Conclusion

A mobile app monitoring inactivity and providing a real-time notification when inactivity period exceeds healthy limits was demonstrated to significantly reduce inactivity periods in overweight sedentary women. The proposed approach is warranted for further investigation in larger group of subjects using randomized clinical trial design.

References

[1] Lee C, Joseph L, Colosimo A, et al. Mortality in diabetes compared with previous cardiovascular disease: A gender-specific meta-analysis. Diabetes Metab. 2012;38(5):420-7.

[2] van der Ploeg HP, Chey T, Korda RJ, et al. Sitting time and all-cause mortality risk in 222 497 Australian adults. Arch Intern Med 2012;172:494-500.

[3] Grontved A, Hu FB. Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis. JAMA 2011;305:2448-2455.

[4] Miller TM, Gilligan S, Herlache LL, et al. Sex differences in cardiovascular disease risk and exercise in type 2 diabetes. J Investig Med 2012;60:664-670.

[5] Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43:1334-1359.

[6] Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Circulation 2007;116:1094-1105.

[7] Barron J, Bedra M, Wood J, et al. Exploring three perspectives on feasibility of a patient portal for older adults. Stud Health Technol Inform. 2014;202:181-4.

[8] Finkelstein J, Cabrera MR, Hripcsak G. Web-based monitoring of asthma severity: a new approach to ambulatory management. Proceedings of the 1998 IEEE International Conference on Information Technology Applications in Biomedicine (ITAB 98), 1998:139-143.

[9] Li J, Garcia S, Castro HK, et al. Acceptance and expectations of information technology to support hypertension self-care in African Americans: a qualitative inquiry. AMIA Annu Symp Proc. 2007:1032.

[10]Finkelstein J, Bedra M. Is Internet search better than structured instruction for web-based health education? Stud Health Technol Inform. 2013;190:65-7.

[11]Rodriguez M, Cha E, Finkelstein J. Feasibility of computer-assisted diabetes education in Hispanic seniors. Proceedings of the 2012 Pan American Health Care Exchanges (PAHCE), 2012:76-77.

[12]Finkelstein J, Knight A, Marinopoulos S, et al. Enabling patient-centered care through health information technology, Evid. Rep. Technol. Assess., (Full Rep) 206, pp. 1–1531. June 2012.

[13]Bedra M, McNabney M, Stiassny D, et al. Defining patient-centered characteristics of a telerehabilitation system for patients with COPD. Stud Health Technol Inform. 2013;190:24-6.

[14]Finkelstein J, Wood J. Mobile eLearning platform for interactive patient education. Proceedings of the International Conference on Mobile, Hybrid, and On-line Learning (ELML '09), 2009:23-27.

[15]Kukafka R, Jeong IC, Finkelstein J. Determinants of health behavior choices in patients using computer-mediated decision aid. Stud Health Technol Inform. 2015;208:226-31.

[16]Finkelstein J, Nambu S, Khare R, et al. Introducing Handheld Computing for Interactive Patient Education. AMIA Annual Symposium Proc. 2002:1019.

[17]Bedra M, Wick E, Brotman D, et al. Avatar-based interactive ileostomy education in hospitalized patients. Stud Health Technol Inform. 2013;190:83-5.

[18]Skinner C, Finkelstein J. Review of mobile phone use in preventive medicine and disease management. Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies, 2008:180-189.

[19]Nyun MT, Aronovitz JR, Khare R, et al. Feasibility of a palmtop-based interactive education to promote patient safety. AMIA Annu Symp Proc. 2003:955.

[20]Finkelstein J, Wood J. Information technology for continuous patient health education. Proceedings of the 2011 conference on information technology education (SIGITE '11), 2011:245-250.

[21]Kukafka R, Jeong IC, Finkelstein J. Introducing decision support for smart mobile health behavior change applications. Proceedings of the 2015 International Conference on Big Data and Smart Computing (BigComp), 2015:75-78.

Address for correspondence

Joseph Finkelstein, MD, PhD, Chronic Disease Informatics Program, Division of Geriatric Medicine and Gerontology, 2024 E Monument St, Baltimore, MD 21205, USA; E-mail jfinkel9@jhmi.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-93

iDECIDE: A Mobile Application for Insulin Dosing Using an Evidence Based Equation to Account for Patient Preferences

Buffy Lloyd[a,b], Danielle Groat[a,b], Curtiss B. Cook[a,b], David Kaufman[a], Adela Grando[a,b]

[a] *Arizona State University Department of Biomedical Informatics, AZ*
[b] *Mayo Clinic Arizona, Division of Endocrinology, Scottsdale, AZ*

Abstract

Diabetes is a complex disease affecting 29.1 million (9.3%) US citizens [1]. It is a chronic illness that needs continual medical care and ongoing patient self-management, education, and support [2]. There is no cure for diabetes, requiring patients to conduct frequent self-monitoring of blood glucose and dosing of insulin in many cases. Evidence has shown that patients are more adherent to their diabetes management plan when they incorporate personal lifestyle choices [3]. To address the challenge of empowering patients to better manage their diabetes, we have developed a novel mobile application prototype, iDECIDE, that refines rapid-acting insulin dose calculations by incorporating two important patient variables in addition to carbohydrates consumed that are not a part of standard insulin dose calculation algorithms: exercise and alcohol intake [4, 5]. A retrospective analysis for the calibration and evaluation of iDECIDE is underway by comparing recommendations made by the application against dosing recommendations made by insulin pumps.

Keywords:

Diabetes mellitus; Insulin dosing; Clinical decision support systems Mobile application Disease self-management.

Introduction

Patient-centered care is defined as health care that respects patients' wants, needs, and preferences, supporting patient desires to make decisions and participate in their own care [6]. Too often patients must adapt to pre-existing protocols and guidelines, rather than receiving services designed to focus on their individual needs and preferences [6]. Patient-centered decision support that translates evidence-based care into health care practice in ways that account for individual preferences and goals is needed.

Many patients with chronic conditions such as diabetes can benefit greatly from self-management [7]. Self-monitoring of blood glucose can be empowering for patients, but tracking such data can be overwhelming [8]. Even patients well trained in diabetes self-management often fail to meet personal glycemic goals. Despite ongoing research to identify patient preferences, track treatments, and integrate patient data to provide personalized options, significant advances in the design and deployment of patient-centered decision aids are still to be made [9, 10].

Type 1 diabetes (T1D) is a chronic disease in which a person's pancreas does not produce insulin, a hormone required to regulate carbohydrate and fat metabolism in the body. Type 2 diabetes (T2D) results from a relative insulin deficit and can be due to a diminished insulin effect or insufficient production to maintain normal blood glucose levels. T2D patients may need insulin injections, oral medications, non-insulin injectable medications, or various combinations of these to control hyperglycemia. Patients with T1D must manage their disease by using insulin injections deliverable through syringes, insulin pens, or insulin pumps. Contemporary insulin pumps utilize a rapid acting insulin analog and deliver continuous basal insulin. Additionally, insulin pumps have bolus calculators that calculate the units of insulin needed based on settings, food intake and active insulin time. Such bolus calculators, which are designed to cover mealtime glucose excursions, do not take into account patient preferences such as alcohol intake and exercise. Evidence shows that these personal preferences can have a significant short-term impact on glucose levels, which in turn affects insulin dosing [4, 5]. Our hypothesis is that by incorporating current evidence regarding the impact of exercise and alcohol intake on insulin dosage, we can further improve postprandial glucose levels for adult individuals with diabetes, thereby empowering them to make informed, evidence-based self-management decisions. We have designed and seek to evaluate a novel, evidence-based decision support tool, iDECIDE, which customizes and refines rapid-acting insulin dosing calculations by incorporating individual preferences for exercise and alcohol. The target population of iDECIDE is adult diabetes patients with T1D or T2D.

Most of the available mobile applications are not evidence-based [11], while iDECIDE is based on the most current medical evidence.

Methods

A literature search that included diabetes pathophysiology, treatment and management options was conducted. We identified insulin dosage calculations based on glycemic levels, carbohydrate intake, exercise, and alcohol consumption.

Next, we reviewed the literature on smartphone apps for diabetes self-management and apps for healthy eating, physical activity, and personal health and wellness [11, 12]. Based on the review there is a proliferation of apps that are not evidence-based or do not align with well-established behavior change theories.

Following our literature review, we met with an endocrinologist and diabetes care team to further understand diabetes and to discuss current clinical challenges that patients encounter. We participated in a guided simulation training session with a diabetes nurse educator at the Mayo Clinic Arizona Simulation Center that included hands-on experience with insulin pumps and continuous glucose monitors. The training excluded review of existing smartphone apps for diabetes management, fitness or nutrition. Based on the trainings we created three prototypical patient cases to reflect the daily regimens and personal preferences encountered on a daily basis by diabetes patients. We learned that diabetes is not a "one size fits all" disease and

that personal management requires special consideration for each patient.

We also reviewed existing insulin pump technologies commercially available in the US. State-of-the-art insulin pumps compute mealtime insulin doses based on proprietary formulas that are approved by regulatory entities like the US Federal Food and Drug Administration (FDA). While alcohol intake and exercise can have an impact on blood glucose levels, no insulin pump takes into consideration alcohol and exercise to compute the insulin needed to correct for a meal. While insulin pumps provide bolus wizards to compute pre-meal insulin boluses, diabetes patients can manually compute pre-meal insulin bolus using an equation from Colin et al. (Equation 1) which takes into consideration important factors, except alcohol and exercise, for choosing the correct insulin dose [13].

$$U = \frac{carbs}{ICR} + \frac{cBG - tBG}{CF} - IOB \quad (1)$$

In Equation 1, the variable *U* represents units of insulin. The first fraction in the equation, "*carbs/ICR*", calculates the relationship between the grams of carbohydrates (*carbs*) intended to be consumed covered by one (1) unit of insulin (*ICR*). *ICR* is calculated as 450/TDD, where Total Daily Dose of insulin (TDD) = body weight (lbs) x 0.23. The second fraction in the equation calculates the difference between the actual blood glucose level (*cBG*) and the target blood glucose level (*tBG*) and divides this difference by the Correction Factor (*CF*). The correction factor, also called insulin sensitivity factor (ISF) is defined as how much one (1) unit of rapid acting insulin lowers an individual's blood glucose over the course of 2-4 hours during a fasting or pre-meal state. These correction doses can account for approximately 9% of the TDD by compensating for the deficits in basal rates or carbohydrate boluses. *CF* is calculated as (1700mg/dl) / TDD. The final segment of the equation subtracts the Insulin On Board (*IOB*) i.e. the theoretical amount of insulin remaining in the body after the last bolus dose.

The ADA states that regular physical activity is important for maintaining health and fitness for those diagnosed with diabetes. People with diabetes are advised to participate in at least 150 minutes of moderate-intensity physical activity per week. Regular exercise has been shown to improve blood glucose control, reduce cardiovascular risk factors, contribute to weight loss and improve well-being [14]. Evidence suggests that most forms of low-to-moderate intensity physical activity result in an increase of insulin sensitivity, which produces a drop in blood glucose levels. When glucose levels drop to abnormally low levels it is called hypoglycemia. Hypoglycemia can be averted by reducing the bolus insulin, increasing food intake, or a combination of both [5]. The evidence recommends ingestion of carbohydrates (e.g. snacks) before exercising to avoid hypoglycemic events.

Alcoholic beverages present an even more complex insulin dosing challenge. Depending on the specific content of the drink, alcoholic beverages can be a carbohydrate source and/or result in delayed hypoglycemia. It is difficult for patients to factor alcoholic drinks into their insulin dosing calculations. Also, they frequently are not aware that more than 2 alcoholic drinks can increase the probability of hypoglycemia a few hours after alcohol consumption [15].

We therefore propose a new insulin dosing equation (patent pending) that accounts for the intensity and duration of physical exercise as well as the alcohol load and related carbohydrates from alcoholic beverages. We have added to the standard equation (Equation 1) parameters to account for patient preferences for exercise and alcohol consumption. As we noted previously, insulin pump calcuators do not consider exercise when calculating insulin dosage, neither do they factor in the effects of alcohol on insulin sensitivity. iDECIDE incorporates these factors to suggest the dosage of rapid acting insulin and sets an alarm to recommend glucose level monitoring in certain circumstances related to alcohol consumption.

Results

Several prototyping platforms such as WireframeSketcher, POP, and Proto.io™ were compared. Proto.io™ emerged as the best choice due to its drag-and-drop intuitive interface for building interfaces. Figure 1 depicts screenshots of the resulting iDECIDE prototype built with Proto.io™.

To exemplify the use of iDECIDE, Figure 1 demonstrates a T2D patient using the app to decide if insulin should be taken before starting a 30 minute, medium-intensity bike ride. Based on the exercise plan and his current blood glucose level of 150 mg/dl, (cBG=150) iDECIDE recommends no insulin and suggests consuption of an additional 10 g of carbohydrates before starting the exercise to achieve a target glucose level of 130 mg/dl (tBG=130) and to avoid hypoglycemia. iDECIDE is using the ICR=10 and CF=20, based on input from the patient's endocrinologist. The IOB=0.75 because the previous insulin dose was 1.25 units from 2 hours prior [16]. To account for exercise (Ex), 0.25 is subtracted off given the short duration (30 minutes) to be completed [17]. The suggested carbohydrates (10 g) were derived from the evidence regarding the patients weight of 150 pounds and the choice of performing 30 minutes of moderate exercise [5].

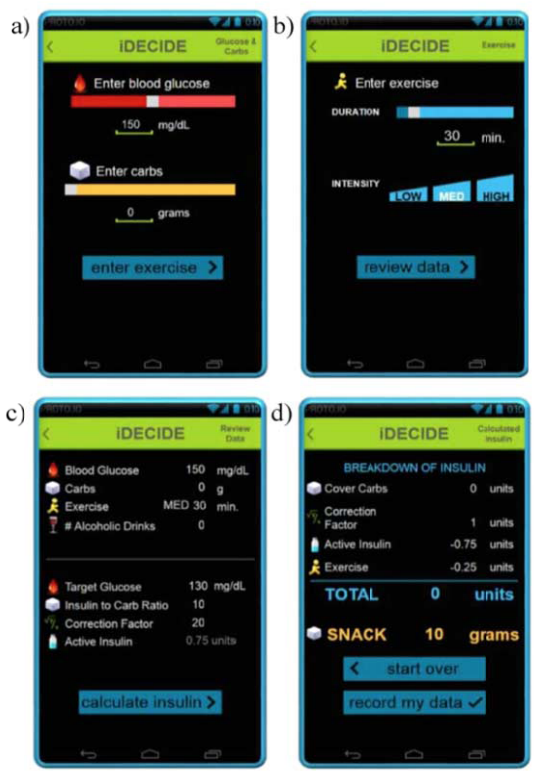


Figure 1- Screenshots of the iDECIDE: a) the patient inputs 150 mg/dl as current blood glucose and that no carbs will be consumed, b) he also inputs that he will be performing 30 minutes of medium intensity exercise; then c) iDECIDE summarizes the input data, the parameters set up by the patient's endocrinologist (e.g. target glucose), and the computed active insulin or IOB; finally d) iDECIDE generates

recommendations (take 0 U of insulin and consume a snack with 10 grams of carbohydrates) and a breakdown of how the suggested insulin dosage was computed (0 U= 0 U to cover carbs + 1 U for correction factor -0.75 U of active insulin – 0.25 U for planned exercise)

Figure 2 exemplifies another use case scenario showing how iDECIDE can set up an alarm if the patient chooses to consume more than 2 alcoholic drinks. The alarm is to remind the patient to monitor blood glucose levels to help avoid hypoglycemic events.

Both Figures 1 and 2 assume that the user enters information immediately before eating, drinking or exercise in order to compute the insulin bolus. It is not uncommon for diabetes patients to input data after they eat or drink to account for last-minute changes.

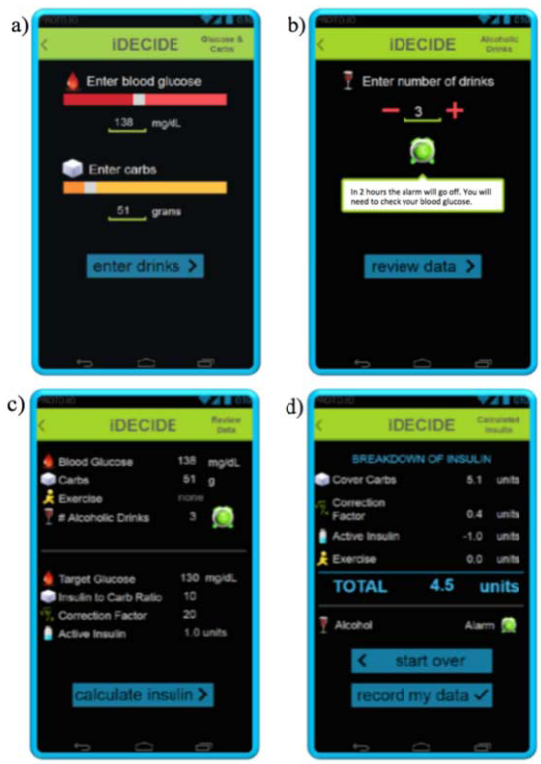


Figure 2- Screenshots of the iDECIDE prototype: a) the patient inputs 138 mg/dl as current blood glucose and that 51 grams of carbs will be consumed, b) he also inputs that he will be drinking 3 alcoholic drinks, to what iDECIDE reacts by setting an alarm to remind checking blood glucose levels to avoid hypoglycemia; then c) iDECIDE summarizes the input data, the parameters set up by the patient's endocrinologist, the computed IOB and indicates that it has setup un alarm; finally d) iDECIDE generates recommendations (take 4.5 U of insulin and check the blood glucose levels in 2 hours when the alarm rings) and a breakdown of how the suggested insulin dosage was computed

We incorporated feedback on this mobile application and on the iDECIDE evidenced-based insulin dosing equation from domain experts in clinical decision support systems and usability, as well as fellow biomedical informatics graduate students. The iDECIDE prototype displayed in Figures 1 and 2 resulted from these recommendations. Then, we deployed the resulting improved interfaces and functionalities as an Android app and we performed a usability study. We secured IRB approval from Arizona State University to recruit 5 students to participate in the study. Participants were given 7 tasks to complete after a 5 minute period of self-guided exploration of the tool. Afterwards they were given a usability survey to complete. A total of 7 usability issues were identified. The exploratory task resulted in the most issues, 5.8, with the final two tasks resulting in no reportable issues. This may suggest that users were able to learn to use the system over time.

The class diagram in Figure 3 shows the main classes (domain concepts) and relationships used for designing iDECIDE. When possible, the domain knowledge of iDECIDE was mapped into terminologies and thesaurus like the Current Procedural Terminology (CPT), the SNOMED Clinical Terms, the National Cancer Institute thesaurus (NCI) and RXNORM. For instance, the concept *currentBG* was mapped to the NCI with the code C0392201.The *Diabetes Patient* using *iDECIDE* takes daily multiple measurements of blood glucose *(currentBG),* which is a type of *Endocrine Finding*. We are also modeling that, for example, a patient can use iDECIDE to set up clinical goals (*hasDesiredState*) related to *Target Glucose, Target Carbs, Target Exercise,* and *Target Alcohol*. For instance, one goal could be to have no more than 2 alcoholic drinks per day during weekends. Every time the patient interacts with iDECIDE he is requested to input his daily preferences (*hasPreferences)* on *Carbs Plan, Alcohol Plan, Exercise Plan,* and *Insulin Plan*. For example, the patient plans to have 3 alcoholic drinks and a dinner, which account for 51 grams of carbs. Based on the input, iDECIDE triggers recommendation messages to remind the patient that his goal was to consume less than 3 drinks, and can also remind the patient the ADA guidelines on alcohol consumption. The patient can decide to follow iDECIDE's recommendations (*makesRecommendations*) or can decide to stick to the original plan, committing to a plan (*hasCommitedTo*). Patients with chronic illnesses, such as diabetes, frequently encounter *Obstacles* when trying to achieve goals or follow treatment recommendations; they are more likely to be successful if back up plans are identified in advance (*hasBackUpState*) and suggested to the patient when obstacles are encountered.

Also, we are incorporating patient Specific, Measurable, Attainable, Realistic and Timely (SMART) goals related to diabetes management, fitness and nutrition to attempt to further *empower patients* to achieve a healthier lifestyle. For instance, "walk more" is too general as a goal. Instead, "I will walk three times a week for 20 minutes" can be measured, is action oriented, can be chosen based on clinician assessment of the patient's clinical state and self-motivation to change behavior, and has a time frame. Patients can understand SMART goals, and the achievement of SMART goals can be assessed and tracked by decision support systems. Therefore, we are currently working on decision mechanisms to provide suggestions to help patients achieve their chosen goals. For instance, in the example case-scenario described above, the patient has a fitness SMART goal of daily lunchtime exercise for 30 minutes at medium intensity. An obstacle arises for the patient: rain. The model incorporates a back-up plan for inclement weather, and suggests an exercise at home (e.g. a 30 minute WiiFit activity) that will achieve his goal. The proposed decision mechanism is inspired by the goal-based clinical decision support planning framework proposed and implemented by Grando, et al. [18, 19] to detect and recover from deviations to standard clinical care plans. In order to specify and reason on SMART goals we have built an ontology using the Ontology Web Language (OWL) using the Protégé tool. Figure 4 depicts a screenshot of the Protégé tool, demonstrating how we model a SMART goal for exercising and the encountered obstacle. The resulting ontology will support the

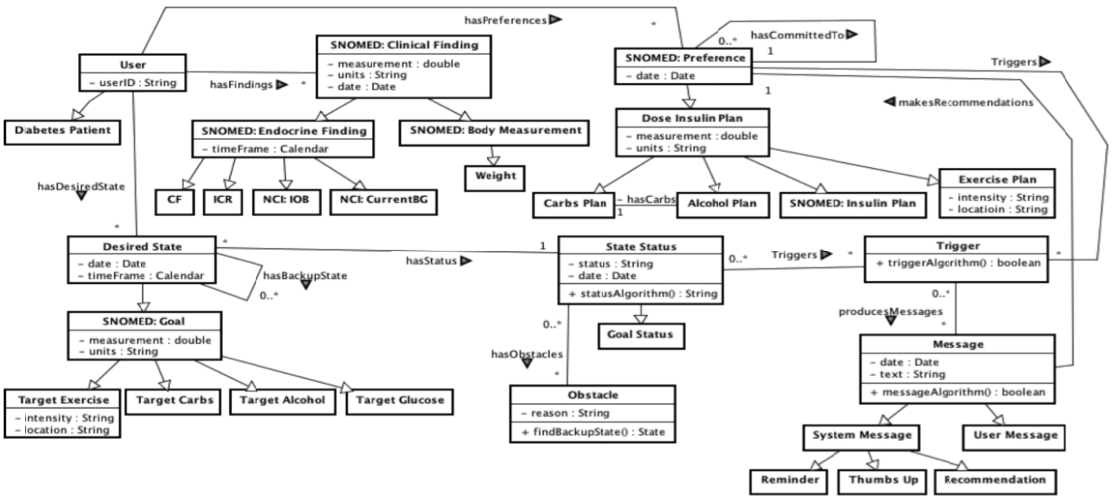


Figure 3- Class diagram depicting iDECIDE's main classes and relationships

decision rules that recommend behavioral changes, such as a specific, pre-identified home exercise option to use when there is inclement weather. Furthermore, using the ontology's goal achievement status (full, partial and none) the achievement status will be automatically determined and tracked. In our example above, the suggested back-up plan, WiiFit, is considered equivalent to the initial outdoor plan, so our patient achieves the prescribed exercise goal.

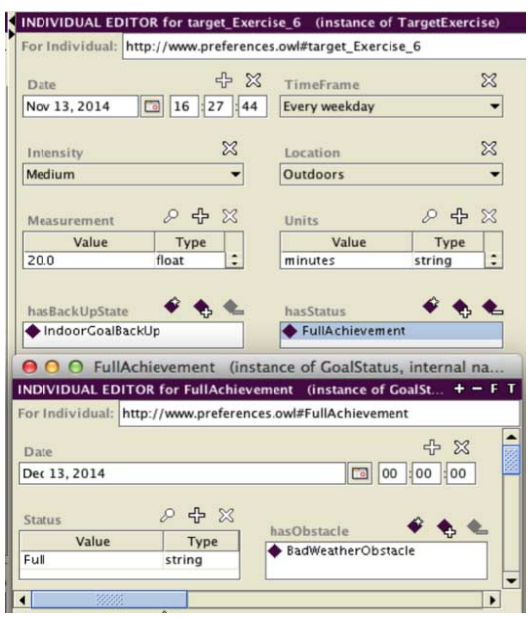


Figure 4- Screenshot from Protégé, displaying the use case scenario of a patient who chooses the goal of walking outdoors every weekday for 30 minutes with medium intensity. In case of in-clement weather, the back-up plan is to exercise at home. The status of the goal achievement (full, partial, none) can be monitored and tracked

Discussion

Future plans include providing reminders, encouragement messages and alternatives to help patients achieve their SMART goals. For our previous example, iDECIDE could remind at lunchtime the patient to exercise with the personalized message, "Time for your lunchtime break exercise. It feels good to be in shape!" The patient can choose to answer the reminder selecting from a set of predefined options including "I cannot exercise today, the weather is bad." Based on the patient's feedback and back up plans iDECIDE can provide suggestions, I see … do you feel like trying some WiiFit tonight instead? I can send you a reminder if you want."

We are recruiting twenty Arizona Mayo Clinic adult patients with T1D who currently use Medtronic™ insulin pumps to begin a retrospective calibration of the evidenced-based formula used by iDECIDE. We are limiting this study to users of Medtronic™ pumps to streamline the data analysis. Participants will be asked to keep records for one month, including alcohol intake and performed exercise (see Tables 1 and 2). The study was approved by the Mayo Clinic IRB.

Initially, we tried to reuse existing retrospective data repositories generated from insulin pumps and patients' diabetes dairies, but the available repositories lacked information on alcohol consumption or they could not be shared due to human subject protection constraints. Following the completion of the study, all participants will provide the data generated during the study period by their insulin pumps and their daily records on alcohol intake and performed exercise. We will input the provided data into iDECIDE. As part of the projected retrospective calibration, domain experts will compare the insulin recommendations generated by iDECIDE against those generated by insulin pumps. We will consider that a recommendation from iDECIDE is as good as the one from the insulin pump when the recommendations are close in range and the postprandial glucose target is achieved. We say that a recommendation from iDECIDE is better than the one from the insulin pump if iDECIDE recommends a higher (lower) dose and the postprandial reading is higher (lower) than target.

We are adopting a user-centered design approach for iDECIDE. Numerous otherwise well-conceived applications that target patients and health consumers have failed to achieve their desired effect because they have not involved users in the development process. Usability issues identified from the completed usability study will be considered to make appropriate design changes. We also plan to conduct another usability study with diabetes patients at the Arizona Mayo Clinic to fur-

ther improve the interfaces and functionalities of the mobile app.

iDECIDE does not communicate with continuous glucose monitors or insulin pumps via wireless or Bluetooth technologies because insulin pump and glucose reader manufacturers do not share the application programming interfaces (APIs) that could facilitate such interactions. iDECIDE currently requires patients to manually input first their glucose reading, meal carbohydrates, alcohol intake and exercise. iDECIDE recommends an insulin dosage to be injected using an insulin pump or syringe. The current study utilizes patients on insulin pumps as a model to test and refine the iDECIDE methodology. Insulin pumps utilize only rapid acting insulin. Future refinement of the system to account for differences in insulin pharmacokinetics will be needed. There are situations where the actions of diabetes patients digress from what was previously entered into the pump's bolus wizard. In these situations there is no technology to account for such behavior. Most patients who use insulin pumps are fairly disciplined and adhere to an established routine, in such cases iDECIDE would be a useful tool.

Table 1– My Diabetic Diary: Tracking alcohol intake

DATE (m/d/y)	TIME (Hour: Min)	TYPE OF DRINK (Beer, wine, etc)	# OF DRINKS	MEASURE (small glass, pint, can, etc)	Did You Input Drink's Carbs Into Insulin Pump?	
__/_ /_	__:__				YES carbs:	NO

Table 2– My Diabetic Diary: Tracking exercise performed

DATE (m/d/y)	TIME (Hour: Min)	INTENSITTY – check one			DURATION (minutes)
__/_ /_	__ : __	LIGHT	MODERATE	VIGOROUS	

Conclusion

iDECIDE is a novel mobile application prototype that personalizes insulin dose calculations by incorporating two important patient variables that are not currently a part of standard insulin dose calculation algorithms: exercise and alcohol intake. Unlike the proprietary algorithms currently employed by insulin pump manufacturers to calculate insulin dose recommendations, iDECIDE is based on available clinical evidence that can be reviewed and discussed by the patient with the endocrinologist and care team. Also, iDECIDE will empower patients to improve disease management, fitness and nutrition by incorporating SMART goals. The app will help to track the achievement of SMART goals, but also provide reminders, encouragement messages and alternatives to help patients achieve their goals.

Acknowledgements

This research was supported by iDECIDE: Patient-centered decision support based on device data (1U54HL108460), funded by NIH National Library of Medicine. We would like to thank Jelena Mirkovic, PhD, for her advice on implementation options. We are grateful to Anita Murcko, MD, FACP for her clinical advice and to Marilyn Bailey, RN for providing training on insulin pumps.

References

[1] National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation. National Diabetes Statistics Report, 2014.

[2] American Diabetes Association. Standards of Medical Care in Diabetes—2011. Diabetes Care 2011: 34(1 Suppl): S11–S61.

[3] V. M. Montori, A. Gafni, and C. Charles. A shared treatment decision-making approach between patients with chronic conditions and their clinicians: the case of diabetes. Health Expect 2006: 9(1): 25–36.

[4] P. J. Phillips, M. Carapetis, and C. Stanton. Alcohol and diabetes. [Online]. Available: http://www.medicinetoday.com.au/2008/june/handout/alcohol-and-diabetes#.VGPZTUtREm8. [Accessed: 12-Nov-2014].

[5] G. Scheiner, R. J. Sobel, D. E. Smith, A. J. Pick, D. Kruger, J. King, and K. Green. Insulin pump therapy: guidelines for successful outcomes. Diabetes Educ 2009: 35(2 Suppl): 29S–41S; quiz 28S, 42S–43S.

[6] M. P. Hurtado, E. K. Swift, and J. M. Corrigan. Envisioning the national health care quality report. National Academies Press, 2001.

[7] R. E. Glasgow, M. M. Funnell, A. E. Bonomi, C. Davis, V. Beckham, and E. H. Wagner. Self-management aspects of the improving chronic illness care breakthrough series: implementation with diabetes and heart failure teams. Ann Behav Med 2002: 24(2): 80–7.

[8] A. M. O'Connor, C. L. Bennett, D. Stacey, M. Barry, N. F. Col, K. B. Eden, V. A. Entwistle, V. Fiset, M. Holmes-Rovner, S. Khangura, H. Llewellyn-Thomas, and D. Rovner. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2009: 3: CD001431.

[9] M. Krahn, and G. Naglie. The next step in guideline development: incorporating patient preferences. JAMA 2008: 300(4): 436–8.

[10] T. E. Quill, and R. G. Holloway. Evidence, preferences, recommendations--finding the right balance in patient care. N. Engl. J. Med. 2012: 366(18): 1653–5.

[11] A. Bastawrous, and M. J. Armstrong. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med 2013: 106(4):130–42.

[12] J. H. West, P. C. Hall, C. L. Hanson, M. D. Barnes, C. Giraud-Carrier, and J. Barrett. There's an App for That: Content Analysis of Paid Health and Fitness Apps. Journal of Medical Internet Research 2012: 14(3): e72.

[13] I. M. Colin and I. Paris. Glucose Meters with Built-In Automated Bolus Calculator: Gadget or Real Value for Insulin-Treated Diabetic Patients?. Diabetes Ther 2013: 4(1): 1–11.

[14] American Diabetes Association. Fitness. http://www.diabetes.org/food-and-fitness/fitness/ Accessed: 01/29/2014.

[15] American Diabetes Association. Alcohol. http://www.diabetes.org/food-and-fitness/food/what-can-i-eat/making-healthy-food-choices/alcohol.html. Accessed: 11/25/2014.

[16] J. Walsh, R. Roberts, and T. Bailey. Guidelines for optimal bolus calculator settings in adults. J Diabetes Sci Technol 2011: 5(1): 129–35.

[17] K. M. Bolderman, K. M. Bolderman, and American Diabetes Association. Putting your patients on the pump. Alexandria, Va.: American Diabetes Association, 2013.

[18] M. A. Grando and A. Boxwala. An ontology-based engine for enacting clinical guidelines in normal and abnormal scenarios. Division Biomedical Informatics Report, DBMI-01-15-2012-001, 2012.

Address for correspondence

Maria Adela Grando, PhD, agrando@asu.edu
Arizona State University, Department of Biomedical Informatics, Mayo Clinic, Samuel C. Johnson Research Building, 13212 East Shea Boulevard Scottsdale, AZ 85259

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-98

Reconfigurable Embedded System for Electrocardiogram Acquisition

Marcel Seiji Kay[a], Fábio Iaione[a]

[a] *Faculty of Computing, Federal University of Mato Grosso do Sul, Brazil*

Abstract

Smartphones include features that offers the chance to develop mobile systems in medical field, resulting in an area called mobile-health. One of the most common medical examinations is the electrocardiogram (ECG), which allows the diagnosis of various heart diseases, leading to preventative measures and preventing more serious problems. The objective of this study was to develop a wireless reconfigurable embedded system using a FPAA (Field Programmable Analog Array), for the acquisition of ECG signals, and an application showing and storing these signals on Android smartphones. The application also performs the partial FPAA reconfiguration in real time (adjustable gain). Previous studies using FPAA usually use the development boards provided by the manufacturer (high cost), do not allow the reconfiguration in real time, use no smartphone and communicate via cables. The parameters tested in the acquisition circuit and the quality of ECGs registered in an individual were satisfactory.

Keywords:
M-health, Electrocardiogram, Embedded systems, Field Programmable Analog Array.

Introduction

The smartphones popularization and the growth of available resources on these devices, as well as higher life expectancy, resulted in the development of mobile systems to be used in the medical field, resulting in an area called Mobile Health (m-Health). Nowadays, this area has generated significant breakthroughs and contributions in health-related research [1]. Studies undertaken by the World Health Organization (WHO) estimates that in 2030, about 23.3 million people will be fatal victims of cardiovascular disease [2]. One possibile solution to reducing these estimates is to increase the use of portable electronic systems to diagnose various heart diseases. These systems easily and comfortably perform tests with mobility, enabling preventative care and avoiding more serious problems [3].

Electronic systems that perform the acquisition of the electrocardiogram (ECG) consist of an analog stage (signal amplification and filtering) and a digital stage (signal digitizing). It's important to emphasize that for the acquisition of bioelectric signals such as ECG, an analog stage will always be necessary due to the necessity of amplifying and filtering the signal before its digitization, eliminating the electrode half-cell potential and avoiding aliasing. Therefore, the Field Programmable Analog Array (FPAA) is an analog stage option. The FPAA is a semiconductor device consisting of several internal configurable analog circuits (switched-capacitor filters, multiplexers, operational amplifiers, and others). The use of FPAA offers the following advantages: easy and fast prototyping circuits, low cost, reduction of the dimensions of the printed circuit board, increased system reliability, high precision and flexibility [4].

There are a few studies designed to use FPAA for the acquisition of ECG signals [3, 5, 6, 7]. In was observed the following limitations in common in these studies: a desktop computer used to view and store ECG signals in real time; usually require a Field Programmable Gate Array (FPGA) and a FPAA contained in their development boards provided by the manufacturer (high cost); the FPAA settings are pre-generated on a computer and stored in the FPGA or an auxiliary memory; the FPAA sends the ECG signal to the computer via cable (USB or EIA-232); and the recording tests are performed using only ECG simulators.

The lack of FPAA studies in literature added to the growing demand for this subject, motivated the goal of this paper to develop a Reconfigurable Embedded System for ECG Signal Acquisition (RES-ESA), and a software to display and store these signals on a smartphone with the operational system (OS) Android. Thus, we seek to explore the advantages provided by FPAA outdoing the limitations of other papers.

Materials and Methods

The microcontroller (MCU) selected was the C8051F320 [8] which is an 8-bit device with a core compatible with the MCS-51 family instruction set. This core has a pipeline that allows the execution of 70% of instructions in 1 or 2 clock cycles, providing an acceleration of approximately 12 times greater compared to the original MCS-51 core using the same clock. In addition to the resources available in the MCUs from the family MCS-51, the C8051F320 also includes: 24.5 MHz oscillator, 2304 bytes data memory (RAM) (256 bytes original, 1 extra kB and 1 kB USB FIFO), 16 kB program memory (FLASH) programmable in the system, SPI serial port, USB 2.0 transceiver/controller, A/D converter (10-bit SAR, 200 ksps, 17 multiplexed inputs), 2.44 V voltage reference to the A/D converter, 3.3 V voltage regulator, crossbar that allows to configure each MCU I/O pin function, among others. The MCU package is a kind of 32-pin SMD (LQFP) and was welded in a DIL adapter card (2.54 mm pitch) allowing its use in the breadboard.

The FPAA used in the study was the AN221E04 [4], which internally has: four configurable analog blocks (CAB), six E/S configurable cells (two output specific), 256 byte Look up table (LUT), voltage references, oscillator and configurable frequency dividers, configuration interface, configuration RAM memory (stores the current FPAA configuration), shadow RAM memory (stores the new FPAA configuration), and analog switches (connect any CAB in any I/O cell). There are internally in each CAB the following analog devices: two operational amplifiers, a voltage comparator and eight capacitors, in addition to analog switches that connect these devices. The two RAM memories allows on-the-fly reconfiguration, in which a new configuration may be sent to the FPAA while the current is in operation, and the new one takes place with no FPAA restart required. The AN221E04 has a 44-pin SMD

(QFP) package and was also welded into an adapter to be used in the breadboard.

The FPAA configuration was performed using the AnadigmDesigner2 software tool that provides the analog circuits implemention graphically generating the data file configurations as needed. The circuit creation used the denominated CAM modules, which can implement analog functions such as filters, amplifiers, multipliers, adders, peak detectors, comparators, rectifiers, among others.

The AN221E04 supports two methods to dynamically change the analog circuits configured parameters (gain, cut-off frequency, among others): the algorithmic method and the state-driven method. The algorithmic method uses functions in C (previously generated by AnadigmDesigner2) to create the FPAA data reconfiguration. The state-driven method, previously is generated by the AnadigmDesigner2 a primary configuration file and secondary configuration files. The latter serve to reconfigure only the parameters of the analog circuit defined by the primary configuration.

The wireless communication was implemented using HC-05 module, which is compatible with the 2.0 Bluetooth standard. The MCU and module communication uses a serial interface EIA-232 (TTL levels), set to 9600 bits/s, eight data bits, no parity bit and a stop bit. The module has an integrated antenna and requires a voltage from 3.3V to 6V.

The Smartphone used was the Samsung Galaxy Nexus I9250, which offers the following specifications: dual-core ARM Cortex-A9 CPU with 1.2 GHz clock, 1GB of RAM, 16GB of flash memory, 3.0 Bluetooth compatible with previous versions, Android 4.2.1, among others.

The integrated development environment (IDE) Silicon Laboratories 4.50 was used in MCU programming in C, installed on Windows 8. This IDE allows creating MCU programs for the MCS-51 family using the open-source Small Device C Compiler [9]. The Smartphone application program was held at Juno Eclipse IDE using the Android SDK (Software Development Kit) 4.2.2 (API 17) installed. It is worth highlighting that the Google USB Driver package (Galaxy Nexus compatible) was installed on Eclipse. This package allows applications to be tested directly on the Smartphone via USB cable. It allows the tests using Bluetooth, because the simulator does not originally support this resource.

Figure 1 shows the system constructed block diagram for the ECG signals acquisition, which was first mounted on a breadboard. It can be observed that the system receives the individual signals by 3 electrodes, whereas all the electronic circuit required for conditioning the ECG signal was constructed in FPAA. Through the SPI bus, the MCU performs the FPAA configuration (MCU acts as the master device and FPAA as a slave device), from the data configuration received from the smartphone via Bluetooth. It is worth pointing that the clock signal (100 kHz) the FPAA requires to work was generated by the MCU, using the programmable couter array (frequency output mode).

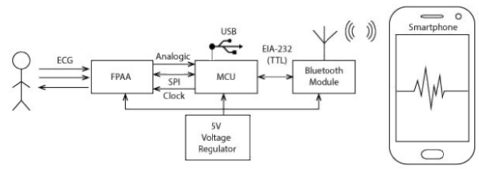


Figure 1 – System block diagram for acquiring ECG signals.

The FPAA amplified and filtered signal is digitized by the MCU A/D converter (8 bits, 480 samples/s, V_{FS} = 3.3V) which transmits the signal to the smartphone via the Bluetooth module. The system works using a 5V source.

The firmware transfer to the MCU used a bootloader, which occupies 5 kB MCU FLASH memory and allows the firmware upload using the computer's USB port and the USBbootloader application installed on an IBM-PC (Windows 8), both provided by the MCU manufacturer. The bootloader firmware was previously recorded on the MCU using the serial EC2 adapter/recorder.

Figure 2 shows the CAMs used and the FPAA internal configuration performed in the AnadigmDesigner2 software.

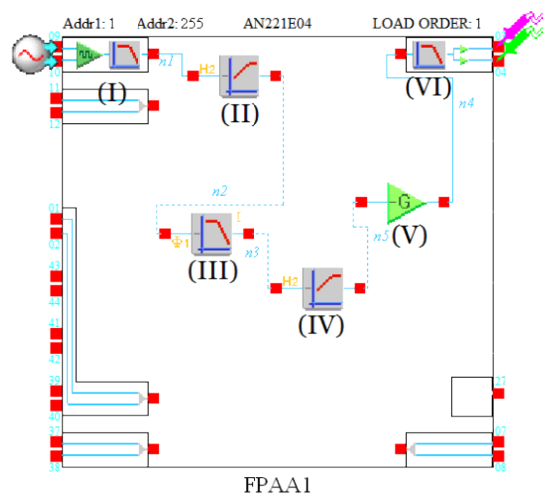


Figure 2 – (I) Input Cell: differential amplifier with low offset (G = 16), bilinear low-pass filter (Fc = 76 kHz); (II and IV) bilinear filter: high-pass filter (Fc = 0.781 Hz), G = 1; (III) biquadratic filter: low-pass filter inverter (Fc = 40 Hz), G = 32.5; (V) Inverter Amplifier: G = -1; (VI) Output Cell: differential output and bilinear low-pass filter (Fc = 76 kHz).

The CAMs (II) and (IV) function is to eliminate the electrodes half-cell potential and OPAMPs offset voltages. The CAM (III) function is to act as anti-aliasing filter. The CAMs (I) and (VI) cut-off frequency was set to the lowest value allowed by the FPAA (76 kHz). The gain is provided by CAMs (I), (III) and (V), the last one can be reconfigured by the smartphone to change the total system gain.

Figure 3 illustrates the developed Android application flow chart.

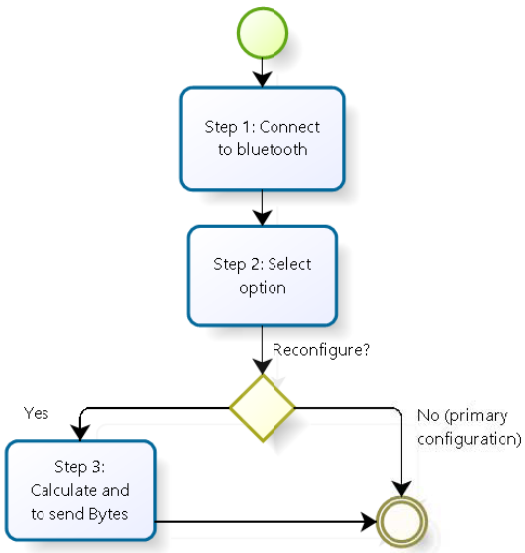


Figure 3 - Initially (step 1), RES-ESA bluetooth pairing. Step 2, the user select on the application if desires to use the primary configuration or reconfigure the FPAA gain. If the reconfiguration is chosen, an algorithm calculates and sends the bytes that change the gain setting of the FPAA. Subsequently, the received data is plotted in real time on the screen.

Results and Discussion

The circuit tests conducted to acquire ECG signals were based on the ANSI/AAMI EC13:2002 [10] recommendations and used the test circuits shown in Figures 4 and 5. The respective tests and results are described below.

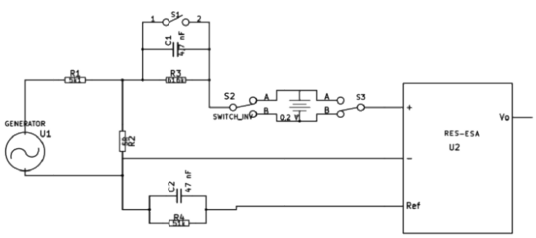


Figure 4 - Diagrammatic model of one of the test circuits used (taken from [10]). The signal from the generator passes through a resistive voltage divider that reduces the amplitude in approximately 1000 times. R4 and C2 simulate the impedance skin-electrode. R3 and C1 are used to estimate circuit input impedance. The 0.2V voltage was generated using a CR2032 battery (3.0 V) and a resistive voltage divider.

Input range and half-cell potential rejection

In this test, the standard specifies that a device must be capable to respond and display voltages ranging from -5 mV to +5 mV. In addition, a +200 mV and a -200 mV tension was injected (simulating the half-cell potential - HCP) in series with the ± 5 mV signal, and a variation in the recorded signal amplitude was registered, which should vary no more than ± 10%. In this test, the signal generator was set to generate a 16 Hz sinusoidal wave and the following settings were used on the Figure 4 circuit, S1 always closed: S2 and S3 in position A (no HCP); S2 in A and S3 in B (with HCP); and S2 in B and S3 in A (with HCP and reversed polarity). It should be observed that it was not possible to use the values specified in the standard to simulate the HCP (± 300 mV), because they saturate the output of the input cell, as its minimum gain is 16 and the output swing is ± 4 V. This was a limitation of the input cell used; however, good Ag-AgCl electrodes present potential around the value used in the test.

Using the data obtained, there were no recorded signal distortions/saturation when the applied sinusoidal wave amplitude was 5 mVp. Furthermore, with HCP, the recorded signal amplitude did not overstepped the acceptable limit of ± 10%. The records average difference with and with no HCP was 0.41 ± 0.19% (α = 99%).

Frequency Response

The frequency response test was performed using two methods, as shown in [10]. In method A, 1 mVpp sinusoid with frequency of 0.67 Hz, 5 Hz, 10 Hz, 20 Hz and 40 Hz were used. The standard requires that the amplitude of the recorded sinusoids come from 0.7A5 to 1.1A5 (A5 is the amplitude of the 5 Hz sinusoidal wave). It was observed, that the amplitudes of the signals recorded in 30 repetitions reached 1.015A5; however, the lower limit was exceeded by the 0.67 Hz frequency reaching 0.63A5, caused by the fact that the high-pass filter has a Fc = 0.781 Hz (minimum possible value), that is, a limitation of FPAA.

Method B used a triangular 1.5 mVpp and 1 Hz signal. For every 10 consecutive cycles with triangle base width of 200 ms, it generated 10 cycles with base width of 20 ms. The standard requires that the recorded signal, the peaks of the 20 ms base triangles must have a minimum range of 75% of amplitude of the 200 ms base peaks. It was found that the 20 ms triangles peaks did not exceed the acceptable lower limit and yielded an average value of 88.6 ± 1.6% (α = 99%).

Noise

According to the standard, the recorded signal noise measurement must not exceed 30 uVpp, when referenced to the input (RTi). To perform this test it was used in a circuit suggested by [10] in which each of the three module terminals was attached to a parallel RC circuit (51 kohms and 47 nF) and connected to a common point (short circuit). It was recorded 10 segments with 10 seconds duration each, and the noise limit must be respected in at least 9 of the 10 segments.

A noise of 67 ± 3 uVpp (α = 99%) was obtained from the tests. However, it should be noted that the circuit was mounted on a breadboard, which is more susceptible to noises due to the length of the wiring connections. Moreover, it was not carried out any detailed analysis of the internal FPAA configuration with the goal to improve this parameter.

Common Mode Rejection

The common mode rejection ratio (CMRR) test, presented the system's ability to reject signals in common mode, as the 60 Hz power grid noise that reaches the body and appears on both system inputs (input+ and input-). The 60 Hz CMRR measurements were applied and recorded two signals: 1.0 mVpp/60 Hz/ differential mode, using the Figure 4 circuit; and 0.4 Vpp/60 Hz/common mode using the Figure 5 circuit. Each record held lasted 60 seconds, the second one was registered with S1 and S2 closed, S1 opened and S2 closed, and S1 closed and S2 opened, in order to simulate an imbalance in the skin-electrode impedance (Figure 5).

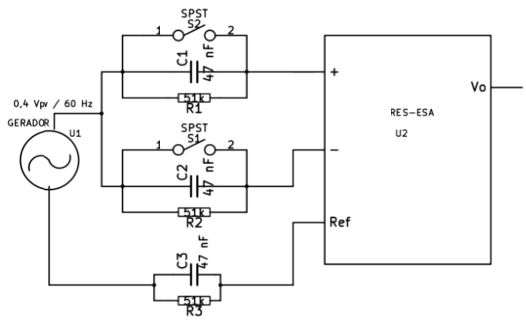


Figure 5 - Schematic diagram of the test circuit used for measuring the CMRR (based on [10]).

Through the analysis of the recorded signals, it obtained the peak to peak voltage for each record and calculated the differential mode gain (G_{md}), common mode (G_{mc}) and the CMRR (CMRR = 20 $\log_{10}$ (G_{md} / G_{mc}). The results were 79.9 dB (S1 and S2 on), 59.4 dB (S1 off, S2 on) and 61.8 dB (S1 on, S2 off).

The standard [10] does not specify values in dB, but the value obtained approximately 60 dB, with unbalance in skin electrode impedance is acceptable, because it indicates that the G_{mc} is 1000 times smaller than G_{md}.

Input Impedance

The input impedance test simulated a skin electrode impedance increase through a 616 kilohms resistor in parallel with a 4.7 nF capacitor connected in series with the positive input (Figure 4). In this test, the recorded signals with S1 opened (Figure 4) should not present an amplitude reduction greater than 20% compared to that obtained without the simulated impedance (S1 closed) within the frequency range from 0.67 to 40 Hz. There was no record that exceeded the 20% limit, with an average reduction of 12.2 ± 2.9% (α = 99%).

Power consumption

The power consumption was measured, disregarding the LEDS electrical currents, using a digital ammeter. The electric current measurement was 88.6 ± 0.6 mA (α = 99%), resulting in an power of 443 ± 3 mW (α = 99%). It should be observed that the FPAA internal circuit configuration optimizations were not performed aiming to reduce consumption. Nevertheless, the consumption obtained was lower in comparison to some other studies using FPAA to acquire bioelectric signals. In study by Mou [3] the system consumption was approximately 12 W (composed of a FPGA development board, a FPAA development board and an LCD screen), whereas in study by Sanches[11], the simulated consumption for electromyogram signal conditioning, FPAA only, was 700 mW. It should be observed that multiple studies [5, 6, 7] did not mention the power consumption.

ECG real signals registration

The test with real ECG signals involved the acquisition of signals in an individual and the DI derivation was used for this (left arm with the positive input and the right arm with the negative input of the system). Figure 6 shows the displayed and stored signal on the Smartphone.

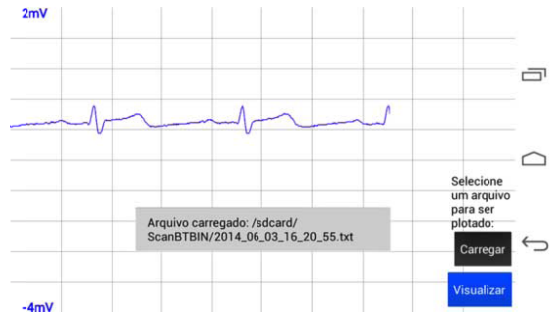


Figure 6 - Real ECG signal recorded in an individual and displayed on the smartphone. This record used the minimum configurable FPAA gain (Total gain = 260).

The signals displayed on the smartphone had a satisfactory reproduction of the real signal captured in an individual, especially considering that the circuit was mounted on a breadboard, a situation that makes it more susceptible to external electromagnetic noise and those generated by the circuit itself (digital part).

Regarding the use of internal FPAA AN221E04 resources, 27 elements were used (56.3%) out of 48 (5 of 8 opamps, 22 of 32 capacitors, none of the 4 comparators and none of the 4 SARs).

Similar smartphone-based systems

The Table 1 shows several existing contemporary systems, displaying the main features of the other smartphone-based systems and the proposed one (RES-ESA).

Table 1– Main Features of the Systems

Name	Description	Price (USD)
AliveCor ECG [12]	Heart monitor with 1-lead (approved by Food and Drug Administration (FDA))	199.00
ECG Check [13]	1-lead, works only on Apple devices (iPhone 4S or newer, iPad 3 or newer, and iPod 5) and approved by FDA.	129.00
Galaxy S5 [14]	Built-in sensor for the heart rate measurement by photoplethysmography.	600.00
RES-ESA	Reconfigurable, 1-lead and low cost.	22.61

Conclusion

In light of the foregoing, it can be noticed that this work successfully used a reconfigurable analog device (FPAA) to capture low amplitude (about 3 mV) ECG signals. The great flexibility offered by the FPAA can be verified with the possibility to reconfigure the amplifiers gain, filter frequency cut-off, filter approximation function and even the signal conditioning circuit. Despite the FPAA advantages, a few studies using FPAA are found about this subject.

The parameters measured in tests (input range, immunity to half-cell potential, frequency response, noise, CMRR, input impedance and power consumption) and the ECG signals quality recorded in an individual were satisfactory. Note that all measured parameters, including the system noise that did not meet the standard, can be improved by optimizing the configured circuit and mounting the entire system on a printed

circuit board. The board has not been manufactured yet, but the layout has already been prepared.

It is possible to view and store the ECG signals in real time on the smartphone Android application and change the total circuit gain, enabling send the data via the Internet to a doctor to examine them later (telemedicine).

The main contributions of this work was the development of a system for acquisition of ECG signals using a smartphone, the development of a reconfigurable system for the acquisition of analog signals, and the creation of Java methods for the Android application to reconfigure the gain of the FPAA via Bluetooth. Comparing this work with others that used FPAAs for the acquisition of bioelectric signals, it can be noticed the following differences: the system developed displays and stores the signals on a smartphone and not only on a desktop computer; the system developed did not use the development board provided by the FPAA manufacturer (high cost); the developed system allows to set the FPAA gain via smartphone in real time; the system sends the recorded data via bluetooth and not only through cables; and the developed system was tested in an individual in real ECG acquisition. It should be noted that no work found allows to view the signals and reconfigure the FPAA on a smartphone.

Some suggestions for future works are: manufacturing the printed circuit board to improve various system characteristics (noise, CMRR, and others); development of a home care system using RES-ESA; analysis of the possible applications and the potential impact on the care delivery loop; use of signal processing algorithms and artificial intelligence on the Android application for QRS detection and heart problems; development and testing new FPAA configurations for other bioelectrical signals acquisition, such as electroencephalogram (EEG) and electromyogram (EMG); use of algorithms for estimation of sleep stages, epilepsy detection, among others; and the development of human-machine interfaces based on bioelectric signals, an area with a lot of attention nowadays.

Acknowledgments

The authors gratefully acknowledge financial support from the Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and the Universidade Federal de Mato Grosso do Sul (UFMS).

References

[1] Silva BM, Rodrigues JJ. A Novel Cooperation Strategy for Mobile Health Applications. In: IEEE Journal on Selected Areas in Communications/Supplement, Vol. 31, No. 9, September 2013; pp 28-36.

[2] World Health Organization. Cardiovascular disease. Available from <http://www.who.int/cardiovascular_diseases/>. Switzerland: World Health Organization, 2013.

[3] Mou P, Chen C, Pun S, Mak P, Vai M. Portable Intelligent Bioelectric Signals Acquisition System with an Adaptive Frontend Implemented using FPGA and FPAA. In: The 11th World Congress on Medical Physics and Biomedical Engineering (WC2009), September 2009; pp 348-351.

[4] Anadigm Incorporation. AN221E04 Datasheet. Available from < http://www.anadigm.com/_doc/DS030100-U006.pdf>. England: Anadigm Inc., 2012.

[5] Grzechca D, Rutkowski J. The use of FPAA in signal processing laboratory for biomedical engineering students. In: International Conference on Signals and Electronic Systems (ICSES), September 2010; pp 453-456.

[6] Morales DP, García A, Castillo E, Carvajal MA, Banqueri J, Palma AJ. Flexible ECG acquisition system based on analog and digital reconfigurable devices. In: Sensors and Actuators A: Physical, Vol. 165, No. 2, 2011; pp 261-270.

[7] Morales DP, García A, Castillo E, Carvajal MA, Parrilla L, Palma AJ. An application of reconfigurable technologies for non-invasive fetal heart rate extraction. In: Medical Engineering & Physics, Vol. 35, No. 7, 2013; pp 1005-1014.

[8] Silabs. Mixed-Signal, 8-bit and 32-bit Microcontrollers (MCUs). Available from <http://www.silabs.com/products/mcu/Pages/default.aspx>. United States: Silabs, 2012.

[9] Small Device C Compiler. SDCC. Available from <http://sdcc.sourceforge.net/>. England: SDCC., 2014.

[10] American National Standards Institute / Association for the Advancement of Medical Instrumentation. ANSI/AAMI EC13:2002 - Cardiac monitors, heart rate meters, and alarms. Association for the Advancement of Medical Instrumentation, United States, 2002.

[11] Sanches PR, Muller A, Carro L, Susin A, Nohama P. Analog reconfigurable technologies for EMG signal processing. In: Revista Brasileira de Engenharia Biomédica, Vol. 23, April 2007; pp 153-157.

[12] AliveCor, Inc. AliveCor Heart Monitor – The mobile ECG. Available from <http://www.alivecor.com/home>. United States: AliveCor., 2015.

[13] Cardiac Designs LLC. ECG Check. Available from < http://cardiacdesigns.com/>. United States: ECG Check., 2015.

[14] Samsung Electronics. Galaxy S5. Available from < http://www.samsung.com/global/microsite/galaxys5/features.html>. South Korea: Samsung., 2015.

Address for correspondence

Marcel.kay@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-103

Availability Communication: Requirements for an Awareness System to Support Nurses' Handling of Nurse Calls

Joakim Klemets[a,b], Pieter Toussaint[a,c]

[a] *Norwegian Research Centre for Electronic Patient Records, Norwegian University of Science and Technology, Trondheim, Norway*
[b] *Department of Telematics, Norwegian University of Science and Technology, Trondheim, Norway*
[c] *Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway*

Abstract

The recent development of mobile technologies allows nurses to receive different types of requests anywhere. However, the interruptions generated by these devices often presents a challenge for nurses in their daily work in a hospital department. In previous inquires we have investigated nurses' strategies to managing technology-mediated interruptions in the form of nurse calls. This study reports on an effort to co-design a system that supports an important strategy employed by nurses. Through the involvement of domain experts, the study elicits requirements for an awareness system to support nurses' collaborative effort in handling nurse calls.

Keywords:

Awareness; Interruptions; Collaboration; Nurse Calls.

Introduction

Nursing work in a hospital department is characterized by complex cognitive work that involves continuous organisation, prioritisation and decision making [1]. Nurses can have multiple clinical and administrative tasks that simultaneously compete for their attention. To manage these challenges in order to provide high quality care, communication among nurses is fundamental [2, 3].

One important aspect of nurses' work is responding to nurse calls. Nurse call systems provide a means for patients to communicate their need of assistance to a nurse. Regardless of how the nurse calls are communicated to the nurse, they are likely to interrupt an ongoing task. This requires the called nurse to assess her availability for the nurse call and perhaps yet again re-prioritise and re-organise work. Therefore, nurse call systems have often been accused to be a source of interruptions that not always are appreciated, as they often disturb ongoing work [4-6]. Interruptions in these environments are also associated with reduced quality of care [7] and patient safety [8].

The intentions behind a nurse call can vary greatly, from toileting assistance or pain medication to information requests [9]. As most nurse call systems do not provide urgency clues, nurses need to decide, with sometimes limited information, whether they should abort an activity to respond to the nurse call or not. A decision that often is regarded as stressful [4, 6].

To date, limited literature exists that describes nurses' handling of nurse calls. Instead, previous research on nurse call systems investigated aspects such as patient satisfaction [10], reasons behind nurse calls [11], and attempts to limit nurse calls [9, 11]. Through ethnographical fieldwork, we previously investigated nurses' strategies in managing interruptions in the form of nurse calls [6, 12]. As a next step, the aim is to discover how a system could be designed to support these strategies.

Nurses assess their availability towards a nurse call based on a number of contextual factors, including their current activity, the relationship to the calling patient and the patient's condition [6]. A strategy employed by nurses to handle nurse calls is to hand over their patient responsibility in certain situations [12]. However, another important strategy identified in our previous studies is maintaining an awareness of colleagues' activities. The decision of whether to respond to a nurse call is highly influenced by this awareness [6].

Therefore, the main objectives in this study were to (1) co-design a system with end-users that supports nurses in maintaining awareness of colleagues' availability, and (2) to analyse the possible effects the proposed system could have on current work. Based on data from our previous studies, we developed a prototype system. The prototype was used as a trigger for further requirement elicitation and analysis in participatory design workshops [13] to achieve the study objectives.

Background

Awareness

Awareness is a central concept within the field of computer-supported cooperative work (CSCW) that is essential when one aims to support efficient coordination and collaboration among team members [14]. The perhaps most commonly used definition of awareness within CSCW is *"...an understanding of the activities of others, which provides a context for your own activity"* [15]. Schmidt discusses the problem that the notion of awareness is indeed ambiguous and that it needs to be understood as a person's awareness of *something* [16]. One relevant example would be activity awareness that is an essential part of collaboration between medical personnel in a hospital. Here, the awareness of colleagues' activities has direct influence on their own activity [17]. Further, Heath et al. discuss how awareness is achieved and sustained through social interaction that involves participants' sensitive monitoring and displaying of actions and activities [16, 18]. A key aspect is that the monitoring and displaying process can occur simultaneously as its own activity, requiring very little effort [19].

There has been extensive research into how technology could support awareness [14]. Within the healthcare domain, one

example is the AwarePhone approach that seeks to minimize interruptions by mediating context cues about the callee to the caller through the phone's address book [20]. The main focus in their study is how the mediated social awareness can enhance collaboration among health care workers who are not co-located. For example, allowing a nurse to consult a doctor more easily. Our focus, however, is to investigate how such an awareness system could be designed to support the handling of a specific type of interruption, that is nurse calls. Nurse calls differ from information requests in the form of phone calls in that the former is expected to be responded to more or less immediately, while that is not always the case for the latter.

Activity Awareness to Support Nurse Call Handling

A few years ago a new, wireless nurse call system was deployed at a Norwegian university hospital. The system allows patients to issue a nurse call by using a draw-string mounted inside each patient room. The signal is then delivered to nurses on both public displays and on wireless phones carried by the nurses. A call plan allows nurses to configure to which nurse a nurse call should be delivered first. If the primary nurse is not able to respond to the call, it can be forwarded to the next nurse. Nurses usually configure the call plan according to agreed patient responsibility. By doing so nurses' receive nurse calls on the wireless phone from patients they have the primary responsibility for first [12].

Through observations and workshops, we investigated nurses' strategies to handle nurse calls at the hospital where the above system was deployed [6]. The findings report a number of contextual factors that play a role in nurses' decision whether to respond to a nurse call in a busy situation. An important factor, however, is their awareness of colleagues' activities.

Nurses continuously monitor colleagues' activities as well as their own, in order to know who is available. Knowing that someone is available when, for example, situated in a patient room talking to an anxious patient, allows the nurse to remain focused on the task at hand, trusting that a colleague will take care of the call. Similarly, nurses prioritize their own activities based on their awareness of colleagues' activities. For example, if a nurse is aware of that his or her colleagues are busy and unable to respond to nurse calls, the nurse may not undertake an activity that will make him or her unavailable as well. Rather, the nurse chooses to remain available in order to act as a backup in responding to nurse calls for the other nurses [6].

However, it is not always the case that nurses are able to monitor colleagues or display their own activities. Nurses report that sometimes a patient visit can take an unexpected turn. In these cases, when nurses normally have not told anyone about the prolonged stay, it might be that no one is able to respond to nurse calls [6].

Methods

Research Design

To meet the first objective, to co-design a system with domain experts, participatory design workshops were held. To trigger discussion, and to demonstrate to the participants technological possibilities, a prototype was developed [21]. The prototype system allowed nurses to communicate their availability status through the wireless phone, as well as, become aware of colleagues availability when a nurse call was issued.

Mogensen and Trigg discuss how artefacts can be used to foster analysis of both current and future practice in participatory design workshops [13]. Scenarios were designed for the workshop where the prototype was used to address the second objective, to analyse possible effects of the designed technology.

Two workshops were held with nine female student nurses from five different hospital departments. All participants had at least one month of experience working at the hospital and using the wireless nurse call system. The workshops were held in a usability lab furnished as a hospital ward and videotaped to facilitate detailed data analysis afterwards. Each workshop lasted for about three hours. After an initial focus group interview, the workshops also included a role-play session, before allowing participants to reflect on their experience with the prototype in another focus group interview.

The scenarios for the role-play session were based on our previous observational studies of the system at the hospital. The scenarios aimed to replicate tricky situations where the acting nurse had to decide whether or not to abort the current activity in order to respond to a nurse call. Participants were instructed to use the prototype in the scenarios. The methodology applied was also influenced by scenario-based design, often employed in the design of collaborative systems [22], and the use of role-play in the design of mobile systems [23].

Prototype

The prototype was developed using a conceptual user interface design tool. The prototype was deployed on an iPhone and allowed users to interact with the application's different screens and menus. However, as the prototype only implemented the user interface, it was not possible to send messages between the phones. Additionally, the application did not receive nurse calls. Therefore, we configured the phones to display a nurse call before the scenario (Figure 1). Then, during the scenario, one of the researchers would play a sound that indicated that the acting nurse received a nurse call.

Figure 1 - Colleagues' statuses are displayed on the phone when a nurse receives a nurse call

In addition to displaying the room number from where the call was initiated, the prototype also displays the availability status for each colleague in the ward. The nurse can then use this information in deciding whether to reject ("avvis") or accept the nurse call ("ok"). Through the user interface, nurses can set their availability status to: available (green), busy (yellow), or unavailable (red), as illustrated in Figure 2. The idea was that a nurse could use the green indication when not occupied with anything important and is able to respond to nurse calls. Yellow indicates that the nurse is busy, but that the task is not of the highest importance. The red indicator communicates that the nurse is undertaking an important task that the nurse would not like to forsake.

Figure 2 - Nurses can be choose from three different availability statuses

Results

Availability Monitoring

When presented with the prototype, nurses' initial reactions were very positive. A nurse reflected: *"That's clever, then you see whether someone can respond to it or not, or if you have to assess responding to it yourself"*. The nurses explained that in a situation where they are occupied with another task, it relives the decision making process if they know that there are other colleagues who can respond to the call, but also that this enhanced awareness of their colleagues' availability could alter their decision to engage in an activity in order to remain an available resource. A nurse explained: *"if I see that many have set themselves to busy I would have waited, it would have been great to have an overview really"*.

This resonates with nurses' current practice of monitoring their colleagues' activities and availability [6]. The nurses reported trying to keep each other informed about activities they undertake. However, as a nurse stated, it is impossible to know where everyone is all the time. For example, nurses said if they are just bringing something from a different floor they do not tell anyone about it. They also admit that it is more difficult to keep track of colleagues' activities while inside a patient room.

Availability Displaying

Being able to display ones' availability through the system to colleagues in times where it would have otherwise been tricky to do so, was seen as a benefit. According to the nurses, this would allow them a higher degree of focus on the task they are performing. A nurse said: *"It would have been useful to be able to give a notice (.), or to lock ones phone (.), so that when you're in a situation that is difficult to leave (.), that some of your responsibility of responding to nurse calls is taken away if there are others who are able to respond during that time"*.

In the focus group discussions, nurses expressed uncertainty in how to set ones' availability status in different situations. The prototype allowed nurses to set their availability to green, yellow, or red. But, it was far from clear in what situation a nurse's availability would be set, for example, to yellow or red. When discussing how the system should behave when different statuses were set, no nurse wished to make themselves completely unavailable (or unreachable) in any situation. They rather wished to indicate that they are occupied at the moment but not completely block incoming requests. While discussing a scenario where the acting nurse was visiting an anxious patient, a nurse said: *"I would never had set myself as unavailable – that is completely unavailable, no, because if the others are busy with various things that they can't leave, then I'm able to leave a patient who is a bit anxious if I explain myself and ask if I can leave for just two minutes"*. Hence, nurses wished to remain in the "loop" even when occupied to remain aware of what was going on outside. As a nurse told: *"You can't remain an overview [of what is happening] if you're completely unavailable or blocked"*. Especially at night shifts, were there are fewer nurses working, it important that nurses are notified about a nurse call even while "unavailable". A nurse explained: *"during the night for example (.), it isn't possible (.), then you are usually alone at the bed area (.), perhaps together with an assistance nurse"*.

One concern the nurses voiced was the extra work maintaining their status would require. Although they saw the benefit, they were afraid it would not be used or that they would forget to change their statuses. Therefore, they suggested the status of a nurse should be changed automatically whenever he or she enters and leaves a patient room. Further, it was proposed that there should be a timer, which would remind the nurses to change their availability back to green after some time.

Adaptive Notification Profile

In the deployed system at the hospital, when a nurse call is issued, nurses are notified through wall panels and the wireless phone. The wall panel where a nurse has marked his or her presence displays the room number and makes an alarming sound. The wireless phone also plays a ringtone when a nurse call is received. The wireless phone at the hospital is configured so that nurses are not able to turn the volume of the ring signal below a certain level.

In the focus group discussions it was evident that the nurses did not appreciate the rigid notification scheme. Nurses felt the ringing was too excessive, especially inside a patient room. A nurse said: *"Inside the room it is quite quiet so you don't need to have such a severe sound"*. Similarly, a nurse expressed that *"there is no point in having alarms both on the phone and on the wall"* inside a patient room. Nurses said the many alarms made it more difficult to focus on the patient they were caring for.

Instead nurses propose that the notification should be modified according to their availability status. A vibration, they explain, would be enough to make them aware of a nurse call in situations where they, for example, are busy with another patient. However, if a nurse call is not responded to by anyone after some time, nurses suggested the notification could become more persuasive. Another nurse proposed that the time the calling patient had waited for a response could be displayed together with the notification.

System Interaction and Feedback

After having forwarded a nurse call, nurses explained that the calling patients do their thoughts, making them wonder whether the patient received help or not. As this further takes some of their attention away from their current task, they said that they would wish to receive feedback, after having dismissed a nurse call, as to whether someone else responded. A nurse proposed that the room number displayed on the wall panel could change colour when a nurse confirmed that he or she will respond to the call. Another nurse confirmed: *"Yes, a type of confirmation that tells you that the nurse call you forwarded has been responded to"*.

Further, the nurses proposed that feedback should be provided to the patient that someone is on their way. The current system does not provide any information to the patient about whether a nurse has noticed their request.

During the focus group interview the nurses explained that the phones are badly suited for sterile environments and the phones should not be exposed in these situations. A badly designed user interface, which required too much attention, was also mentioned as a reason why the phones were not used. The nurses felt that interacting with the phones took too

much focus away from, for example, a visited patient. When asked how they would like to be notified about a nurse call, a nurse explained: *"To have a look at the wall-panel is not as disruptive as to pick up and look at the phone"*. Instead nurses preferred to use the wall-mounted displays to learn about a nurse call and to respond to them. It was suggested that colleagues' availability status should be displayed on public displays, to make the information accessible "at-a-glance". A nurse said: *"I think that the proposal made was very good, that the sound is turned off, and that you instead see it on the wall-panel"*.

Expected Effects

One expected effect of the new functionality the nurses discussed was that it would allow them to more fully focus on the patient they are nursing if they knew there were others available to care for a calling patient. The nurses explained that if they are aware a colleague is able to respond to a nurse call, they then do not rush away from the patient as quickly as they would otherwise.

Further, they expected the availability awareness to reduce the wait-time for patients. Although the nurses said the current system works as it is, the proposed functionality would make it more effective. A nurse explained: *"With the new functionality it would probably be more efficient and faster, five minutes for us seems like nothing, but for the patients it probably feels like forever"*. If the nurses know their colleagues are unavailable, there is no need to let the nurse call keep calling to find out whether someone is able to respond.

Similarly, nurses revealed that, even if they are not busy with something, they sometime hesitate to respond to a nurse call from a patient that is not their assigned responsibility. A nurse explains: *"It is best for the patient to be [in] contact with the same nurse as much as possible, so if she isn't busy there is no reason for me to go in there, otherwise you could just enter any room at any time"*. However, if the primary nurse for the patient is not responding and cannot be seen, nurses said they do respond to the call. Providing quick access to information about the primary nurse's availability status could therefore also reduce response times to nurse calls.

Summary of Proposed Functionalies

The following list summarizes the additional functionalities to the proposed awarenss system that were suggested by the nurses during the workshops.

- Nurse calls should not be blocked even if status is set to yellow or red so nurses can remain aware of what is going on in the department
- The notification for nurse calls should be modified according to set availability status and for how long the calling patient has waited
- Status should be set automatically when entering or leaving a patient room
- A timer should remind the nurses to update their status if it has been set to yellow or red
- Allow nurses to interact with wall mounted displays instead of the phone (monitoring and displaying availability)

Discussion

A concern raised during the workshop was the extra work required in maintaining ones' status, which also was mentioned as an issue in [20]. While the proposed solution to update the status automatically based on location is appealing, there is reason to be cautious when building context-aware applications [24]. Brown and Randell argue for a defensive use of context that allows users to easily correct the system when it makes an erroneous inference [25]. The system could well infer a nurse's availability based on the location, but make this inference flexible. For example, the system could automatically set the nurses status to busy (yellow) when entering a patient room. At the same time, this selection could be displayed on the wall panel inside the room and allow the nurse to change the status by pressing either a green or red button on the wall panel.

As both monitoring and displaying of co-located colleagues requires barely any effort [18, 19], the difficulty is to achieve the same awareness when designing a system to support spatially distributed team members [14]. Some degree of automatic inference of availability helps at the displayer's side, but it is important to not overload monitor with too much information, as Gross discusses [14]. In their paper, Avrahami et al., demonstrate how people often over- and under-estimate the significance of various cues when estimating the interruptibility of a person. In their study, contextual variables such as whether a person was on the phone, drinking, reading, or socially engaged were included. Similarly, another study found that people use contextual cues to merely find out whether a person is present, rather than assessing whether the person can be interrupted or not [26]. While the cues included in these two studies were more related to typical office work, it does raise the question whether such cues are beneficial when estimating whether a person can be interrupted or not? The main rationale to merely display a colour code is to lessen the effort required by the one monitoring. However, whether such an indication is correctly estimated is not addressed in this study.

With regard to system interaction, the results hint that the nurses do not prefer to pick up the phone when a nurse call is issued. Therefore, one alternative could be to present the number of available nurses along with the nurse call on a wall panel. Further, the availability information could be made easily accessible on a big screen in the hallway. Allowing nurses to change their status on the publicly available screen might also provide advantages over doing so through the wireless phone in form of reduced effort.

Scholars have argued for the need of user-centred design approaches in the development of technological systems within the healthcare domain [27, 28]. Participatory design has previously been used in the design of nursing tools [29], and a similar method was adopted in this study. The approach, which combined artefacts in the form of scenarios and a prototype, was found to stimulate participants to come up with ideas on how to design the system; system features that maybe participants would not have been able to envisage otherwise. The prototype, although an early version, allowed the participating nurses to widen their technological frame of reference [30].

The artefacts also trigged analytical discussions on the expected effects of the proposed awareness system. A short summary of these would include; faster response times to nurse calls, less noise, and allow the nurse to remain more focused on the patient. Both the reduced response time and the notification modification based on status, contributes to less noise in a department. Less noise means that nurses can better focus on their work. Also, knowing that someone else is able to respond to an issued nurse call allows nurses to more fully remain focused on the current task.

A study limitation is that all participants were nurse students with limited experience. However, the initial prototype was designed based on a thorough study of nurses work practices that included both observations and workshops with experienced nurses [6]. Another limitation is the relatively small sample size. Yet, in usability studies, for example, there is no consensus of an optimal number of participants [31]. Some advocate that five is enough, while others suggest around ten participants [32], or even more [31].

Conclusion

Through a particpatory design approach, this study has investigated how an awareness system could be designed to support nurses' handling of nurse calls in a hospital setting. Requirements for a system that is sensitive to and communicates availability information of nurses has been elicted. Further, through co-analysis with domain experts, expected effects of the co-desiged system hints at increased focus on current tasks, reduced noise, and faster response times to nurse calls.

Acknowledgement

The authors would like to thank Monika Grimstad Hafredal and Veronica Sund for their help in planning and conducting the workshops, as well as making the prototype.

References

[1] Ebright PR. The complex work of RNs: Implications for healthy work environments. Online Journal of Issues in Nursing 2010: 15(1).

[2] Miller ET, Deets C, and Miller RV. Nurse call and the work environment: Lessons learned. J Nurs Care Qual 2001: 15(3).

[3] Toussaint PJ, and Coiera E. Supporting communication in health care. Int J Med Inf 2005: 74(10).

[4] Klemets J, and Kristiansen L. Extended communication possibilities for nurses: Taking context into consideration. Stud Health Technol Inform 2013: 194.

[5] Deitrick L, Bokovoy J, Stern G, and Panik A. Dance of the call bells: Using ethnography to evaluate patient satisfaction with quality of care. J Nurs Care Qual 2006: 21(4).

[6] Klemets J, and Evjemo TE. Technology-mediated awareness: Facilitating the handling of (un)wanted interruptions in a hospital setting. Int J Med Inf 2014: 83(9).

[7] McGillis Hall L, Pedersen C, and Fairley L. Losing the moment: Understanding interruptions to nurses' work. J Nurs Adm 2010: 40(4).

[8] Mayo AM, and Duncan D. Nurse perceptions of medication errors: What we need to know for patient safety. J Nurs Care Qual 2004: 19(3).

[9] Torres SM. Rapid-cycle process reduces patient call bell use, improves patient satisfaction, and anticipates patient's needs. J Nurs Adm 2007: 37(11).

[10] Roszell S, Jones CB, and Lynn MR. Call bell requests, call bell response time, and patient satisfaction. J Nurs Care Qual 2009: 24(1).

[11] Meade CM, Bursell AL, and Ketelsen L. Effects of nursing rounds: on patients' call light use, satisfaction, and safety. AJN The American Journal of Nursing 2006: 106(9).

[12] Klemets J, Evjemo TE, and Kristiansen L. Designing for redundancy: Nurses experiences with the wireless nurse call system. Stud Health Technol Inform 2013: 192.

[13] Mogensen P, and Trigg RH. Using Artifacts as Triggers for Participatory Analysis. Proc of PDC'92.

[14] Gross T. Supporting effortless coordination: 25 years of awareness research. Comput Support Coop Work 2013: 22(4-6).

[15] Dourish P, and Bellotti V. Awareness and coordination in shared workspaces. In Proceedings of CSCW'92.

[16] Schmidt K. The problem with 'awareness': Introductory remarks on 'awareness in CSCW'. Comput Support Coop Work 2002: 11(3).

[17] Bardram J, and Hansen T. Context-based workplace awareness. Comput Support Coop Work 2010: 19(2).

[18] Heath C, Svensson MS, Hindmarsh J, Luff P, and vom Lehn D. Configuring awareness. Comput Support Coop Work 2002: 11(3).

[19] Schmidt K. Cooperative Work and Coordinative Practices. Cooperative Work and Coordinative Practices: Springer London, 2011; pp. 3-27.

[20] Bardram JE, and Hansen TR. The AWARE architecture: Supporting context-mediated social awareness in mobile cooperation. In Proceedings of CSCW'04.

[21] Kyng M. Designing for cooperation: cooperating in design. Commun ACM 1991: 34(12).

[22] Bardram J. Scenario-based design of cooperative systems. Group Decis Negot 2000: 9(3).

[23] Svanaes D, and Seland G. Putting the users center stage: Role playing and low-fi prototyping enable end users to design mobile systems. In Proceedings of CHI'04.

[24] Erickson T. Some problems with the notion of context-aware computing. Commun ACM 2002: 45.

[25] Brown B, and Randell R. Building a Context Sensitive Telephone: Some Hopes and Pitfalls for Context Sensitive Computing. Comput Support Coop Work 2004: 13.

[26] Fogarty J, Lai J, and Christensen J. Presence versus availability: the design and evaluation of a context-aware communication client. Int J Hum-Comput St 2004: 61(3).

[27] Ongenae F, Duysburgh P, Verstraete M, Sulmon N, Bleumers L, Jacobs A, et al. User-driven design of a context-aware application: An ambient-intelligent nurse call system. In Proceedings of PervasiveHealth'12.

[28] Moen A, and Mæland Knudsen LM. Nursing Informatics: Decades of Contribution to Health Informatics. Healthcare Informatics Research 2013: 19(2).

[29] Reeder B, Hills RA, Turner AM, and Demiris G. Participatory Design of an Integrated Information System Design to Support Public Health Nurses and Nurse Managers. Public Health Nurs 2014: 31(2).

[30] Orlikowski WJ, and Gash DC. Technological frames: making sense of information technology in organizations. ACM Trans Inf Syst 1994: 12(2).

[31] Schmettow M. Sample size in usability studies. Commun ACM 2012: 55(4).

[32] Hwang W, and Salvendy G. Number of people required for usability evaluation: the 10±2 rule. Commun ACM 2010: 53(5).

Address for correspondence

Joakim Klemets, Norwegian University of Science and Technology, Department of Telematics, O.S. Bragstads plass 2a, N-7491 Trondheim, Norway. Email: joakim@item.ntnu.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-108

Optimizing Decision Support for Tailored Health Behavior Change Applications

Rita Kukafka[a], In cheol Jeong[b], Joseph Finkelstein[b]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*
[b] *Chronic Disease Informatics Program, Johns Hopkins University, Baltimore, MD, USA*

Abstract

The Tailored Lifestyle Change Decision Aid (TLC DA) system was designed to provide support for a person to make an informed choice about which behavior change to work on when multiple unhealthy behaviors are present. TLC DA can be delivered via web, smartphones and tablets. The system collects a significant amount of information that is used to generate tailored messages to consumers to persuade them in certain healthy lifestyles. One limitation is the necessity to collect vast amounts of information from users who manually enter. By identifying an optimal set of self-reported parameters we will be able to minimize the data entry burden of the app users. The study was to identify primary determinants of health behavior choices made by patients after using the system. Using discriminant analysis an optimal set of predictors was identified. The resulting set included smoking status, smoking cessation success estimate, self-efficacy, body mass index and diet status. Predicting smoking cessation choice was the most accurate, followed by weight management. Physical activity and diet choices were better identified in a combined cluster.

Keywords:

Decision Support Systems; Self-Management, Statistical Analysis; Consumer Behaviour.

Introduction

Health behaviors play a significant role in maintaining optimal health and preventing chronic health conditions. Information technology has great potential in promoting healthy behaviors in the general population [1-2] and for people with chronic conditions [3-4]. A variety of mobile applications have been introduced to modify health behaviors [5] including mobile smartphone applications [6] and web-based platforms [7].

Decision-aids designed to incorporate patient preferences into health decisions have been shown to facilitate patient-centered care [1]. We designed the Tailored Lifestyle Change Decision Aid (TLC DA) system to provide support for a person to make an informed choice about which behavior change to work on when multiple unhealthy behaviors are present [8]. TLC DA can be delivered via multiple health communication channels including web, smartphones and tablets. The system collects a significant amount of information, which is used to generate tailored outputs to consumers in order to assist them in choosing a health behavior to modify.

One of the current limitations of the system is the necessity to collect vast amounts of information from users who have to manually enter all required data. The acceptance, scalability, and ease of use of the mobile app can be significantly improved if we were able to minimize the number of self-reported parameters collected from users at the initial introduction of the app. In this project we aimed at identifying the most important parameters defining health behavior choice in users of the TLC DA using discriminant analysis. By identifying an optimal set of self-reported parameters we will be able to minimize the data entry burden of the app users. Decision support based on a model that can forecast health behavior choice can help further minimize the amount of self-reported data collected. Currently, factors that determine behavior choices by TLC DA users are unknown. The main goal of this study was to identify primary determinants of health behavior choices made by patients after using the TLC DA system.

Methods

System Design and Data Collection

TLC integrates clinical data elements (height, weight, waist circumference, total cholesterol, HDL cholesterol, systolic and diastolic blood pressure, diabetes diagnosis) with patient assessment data to run its decision logic for the tailored outputs [9]. Patient assessment consists of five standardized questionnaires pertaining to 1) diet, 2) weight 3) physical activity 4) tobacco use and 5) psychosocial issues. For each of the four behaviors, we also assess the level of engagement in the unhealthy behavior (i.e., number of cigarettes smoked and level of nicotine dependence) as well as the patient's transtheoretical stage of change for improving the behavior and self-efficacy in their ability to sustain the behavior change. The patient output report contains four components: 1) behavioral risk and clinical risk; 2) readiness and confidence scores for changing each of the four behaviors; 3) qualitative equations to elicit patient priorities for change and 4) an action plan to affirm their behavior change priorities and goals. The readiness and confidence of the individual for each behavior is combined to present chance of success for each behavior. This allows the patient to contrast chance of success (readiness + confidence) against the level of health benefit the patient would achieve if they elect to work on that specific behavior. For example, whereas quitting smoking would produce the highest health benefit, the patient would also have a lower chance of success in quitting at this point in time (Figure 1). On the other hand, although physical activity would produce a lower benefit, the patient would also have the highest chance of success of including physical activity into their lives. Patients are asked to choose a behavior to work on as they consider level of risk and chance of success for each of the unhealthy behaviors they engage in.

Eighty consecutive adults (45 females and 35 males) participated in the study. Participants were recruited via telephone and at community centers, fairs, and other public places and were eligible if they where English speaking, between the ages of 18 and 70 and had no counterindications for engaging in a lifestyle modification progam.

Analysis

Baseline patient characteristics were represented by 45 variables reflecting their age, blood pressure, body mass index (BMI), smoking status and behavioral characteristics associated with behaviors of choice.

The study subjects were grouped based on their choice of health behavior after using the decision aid to corresponding categories including "Physical Activity," "Diet," "Weight Management," and "Smoking". Detailed characteristics of these variables were described previously [10]. Stratified analysis based on patient choice has been performed.

All statistical analyses were performed using IBM SPSS Statistics 22. Group statistics were conducted to examine the overall means and standard deviations of 45 independent variables for four patient's behavior categories. Cross-correlations of group variables were investigated by bivariate correlation analyses. Discriminant analysis was conducted 1) to find the best variable set that determines choice of health behavior; 2) to compose discriminant functions based on linear parameter combinations; and 3) to build a predictive model determining most likely choice of health behavior patients decide to adapt after using a computer-mediated decision aid. To achieve this, a stepwise analysis was carried out, specifically Wiks' lambda method was used.

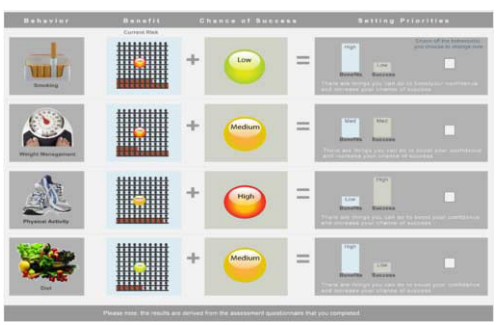


Figure 1 – Tailored Lifestyle Counseling (TLC) System

Results

Group Statistics and Correlations

The means and standard deviations of 45 baseline variables were calculated for each of the four patient behaviors. All nominal variables were converted into ordinal variables for the purpose of this analysis. The study subjects were grouped based on their choice of health behavior after using TLC to corresponding categories including "Physical Activity," "Diet," "Weight Management," and "Smoking". For each group average characteristics were calculated as presented in Table 1.

Bivariate correlations between each "Physical Activity" group variables ranged from -0.236 to 1.000 as can be seen in Table 2. "Diet," "Weight Management," "Smoking" group variables correlations were ranged from -0.620 to 1.000, from 0.290 to 1.000, and from -0.566 to 0.961 respectively as shown in Tables 3, 4 and 5.

Table 1 – Independents Depending on Patient's Behavior

	PA		**Diet**		**WM**		**Smoking**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age	49.72	11.00	52.09	10.22	51.81	12.39	50.09	9.07
SBP	142.72	12.22	140.74	14.29	147.10	17.39	138.09	11.25
DBP	92.24	19.70	88.70	15.88	85.48	10.25	93.64	18.53
BMI	34.65	6.94	31.35	5.57	28.38	5.56	26.79	8.67
Weight	220.40	58.90	193.48	42.28	185.33	43.98	168.73	50.70
Weight Normal	155.11	18.30	151.84	14.90	159.88	18.99	157.35	29.14
WM Readiness	3.84	1.03	4.04	1.07	4.10	1.14	3.82	1.83
WM Confidence	2.08	0.57	2.22	0.74	2.24	0.62	2.09	0.83
WM Risk	9.64	5.52	10.26	6.60	8.67	5.42	6.18	2.23
WM SoC	3.84	1.03	4.04	1.07	4.10	1.14	3.82	1.83
WM Success	2.72	0.54	2.65	0.49	2.76	0.44	2.45	0.69
WM Benefit	2.60	0.50	2.52	0.67	2.38	0.74	2.36	0.50
PA A score	3.36	1.87	2.09	1.16	2.67	1.20	3.00	1.90
PA S score	0.96	1.31	0.87	1.18	0.95	1.24	1.18	1.47
PA SE	15.84	4.76	14.65	4.75	16.86	6.46	17.00	6.87
PA Readiness	3.72	1.43	3.17	1.27	3.76	1.26	3.91	1.64
PA Confidence	1.72	0.54	1.52	0.59	1.90	0.77	1.91	0.83
PA Risk	6.32	3.30	7.83	3.56	7.29	3.30	6.73	4.38
PA SoC	3.72	1.43	3.17	1.27	3.76	1.26	3.91	1.64
PA Success	2.40	0.71	2.04	0.82	2.52	0.68	2.45	0.69
PA Benefit	2.24	0.60	2.43	0.66	2.48	0.60	2.27	0.65
D status	5.36	3.19	4.43	2.81	7.10	2.57	7.00	3.00
D SE Positive	30.32	14.97	36.74	16.90	32.43	13.85	30.36	13.83
D SE Negative	25.60	12.68	32.04	10.73	28.00	11.18	27.45	11.69
D SE Habitual	19.00	7.98	23.13	8.19	20.00	6.75	18.18	9.41
D SE Temptation	39.40	16.80	46.91	16.89	38.00	17.64	36.82	18.29
D Readiness	3.52	1.33	3.91	1.31	2.62	0.80	2.91	1.70
D SE	1.96	0.84	2.43	0.79	2.14	0.85	2.00	0.89
D Risk	6.56	3.36	6.61	2.92	7.90	4.01	5.91	2.43
D SoC	3.52	1.33	3.91	1.31	2.62	0.80	2.91	1.70
D Success	2.40	0.76	2.61	0.58	2.29	0.72	2.00	0.77
D Benefit	2.20	0.65	2.30	0.63	2.43	0.75	2.18	0.60
S Status	0.32	0.90	0.04	0.21	0.29	0.90	4.00	2.14
S SE Positive	6.08	6.81	6.78	6.91	3.33	5.63	8.55	3.14
S SE Negative	5.76	6.83	6.65	6.99	3.43	5.90	5.82	2.27
S SE Habitual	6.56	7.17	7.39	7.39	3.57	6.05	8.27	3.26
S Readiness	4.40	1.41	4.91	0.42	4.76	0.89	2.45	1.21
S Confidence	1.92	1.00	1.91	1.00	1.48	0.87	2.45	0.69
S Risk	7.80	5.13	8.48	5.09	6.76	3.03	9.36	4.32
S SoC	1.80	2.29	2.30	2.49	0.95	1.91	2.00	1.10
S Success	2.32	0.48	2.48	0.51	2.19	0.40	2.45	0.52
S Benefit	2.32	0.56	2.39	0.58	2.33	0.73	2.55	0.52
PS SE Support	3.98	0.91	3.80	0.95	4.09	0.82	3.36	1.09
PS SE Stress	9.00	3.15	9.04	2.85	10.05	2.69	7.82	2.93
PS SE Literacy	2.68	1.73	3.09	1.56	3.29	1.38	2.45	1.69

PA: physical activity, *WM:* weight management, *SD*: standard deviation, *D:* diet, *S:* smoking, *PS*: psycho-social,
N are: 25 for **Physical Activity**, 23 for **Diet**, 21 for **Weight Management**, 11 for **Smoking** and 80 for **Total**.

Canonical Discriminant Functions

To discriminate a priori defined categorical groups ("Physical Activity," "Diet," "Weight Management," and "Smoking"), a stepwise discriminant analysis was performed. At each step, the parameters that minimized the overall Wilks' Lambda were entered. The maximum number of steps was 90, entered

independent variables were S Status, BMI, S Success, S Soc and D Status, and the other 40 independents were eventually removed. In the resulting model, S Status had 0.744 of tolerance, 55.536 of F to Remove and 0.583 of Wilks' Labda, BMI had 0.808 of tolerance, 9.481 of F to Remove and 0.246 of Wilks' Lambda, S Success had 0.069 of tolerance, 9.578 of F to Remove and 0.246 of Wilks' Labda, S SoC had 0.075 of tolerance, 7.188 of F to Remove and 0.229 of Wilks' Lambda, D Status had 0.932 of tolerance, 4.322 of F to Remove and 0.208 of Wilks' Labda, and the model made by these 5 parameters had 11.632 of F to Remove and 0.176 of Wilks' Lambda as shown in Table 6.

As the result of this analysis, three canonical linear discriminant functions were generated to calculate Function 1, Function 2 and Function 3:

Function 1 = 1.060 S Status – 0.085 BMI + 4.779 S Success - 0.893 S SoC + 0.057 D Status – 8.131

Function 2 = 0.288 S Status + 0.094 BMI – 1.335 S Success + 0.488 S SoC - 0.256 D Status + 0.641

Function 3 = 0.145 S Status + 0.095 BMI - 3.184 S Success + 0.437 S SoC + 0.147 D Status + 2.815

The eigenvalues of Function 1, 2 and 3 were 2.989, 0.300 and 0.095 respectively. The proportion of discriminating abilities of the Function 1, 2 and 3 were 88.3%, 8.9% and 2.8% respectively. The canonical correlations of Function 1, Function 2 and 3 were 0.866, 0.480 and 0.295 respectively as shown in Table 7.

Classification Statistics

The classification was processed for all 82 patient variable sets, that included 25 sets of "Physical Activity," 23 sets of "Diet," 21 sets of "Weight Management," and 11 sets of "Smoking," and the prior probabilities were set for all groups equal. As shown in Figure 2, the group centroids (Function 1, Function 2) of "Physical Activity," "Diet," "Weight Management," and "Smoking" were (-0.968, 0.395, 0.346), (-0.726, 0.278, -0.428), (-0.232, -0.889, 0.040) and (4.162, 0.216, 0.031) respectively.

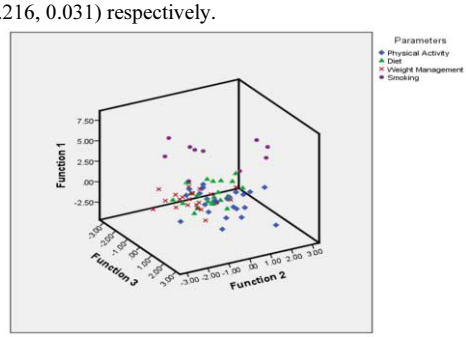


Figure 2- Canonical Discriminant Functions

The classification results are shown in Table 8. The classification success was assessed using original and cross-validated modes in case of our data sets did not differ. Percentage of correct and incorrect classifications was based on the generated Function 1, 2 and 3. Sixty-four percent of "Physical Activity" observations were classified correctly as "Physical Activity", but the remaining cases were incorrectly classified as "Diet" (24%) and "Weight Management" (12%). "Diet" observations were correctly identified 56.5% of the time as "Diet" but 26.1% were misclassified as "Physical Activity" and 17.4% as "Weight Management." "Weight Management" observations were correctly classified 85.7% as "Weight Management" but 4.8% were misidentified as the other behaviors. In cross-validation each case was classified by the functions derived from all cases using the leave-one-out method. There was difference in predicting patient's behavior between the original validation and the cross-validation (Table 8). Original classification estimates yielded overall 71.3% of correctly classified cases using Function 1, 2 and 3 generated by discriminant analysis. Cross-validation resulted in 72.7% correct classification results for the "Smoking" category, 66.7% - for the "Weight Management" category, and 75.5-86.6% - for a combined group of "Physical Activity" and "Diet."

Table 6 – Parameters in the Discriminant Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	S Status	1.000	41.916	
2	S Status	0.988	41.143	0.826
	BMI	0.988	5.154	0.377
3	S Status	0.895	46.927	0.771
	BMI	0.910	6.219	0.333
	S Success	0.854	4.336	0.312
4	S Status	0.744	59.510	0.716
	BMI	0.840	8.544	0.281
	S Success	0.069	9.361	0.288
	S SoC	0.077	6.775	0.266
5	S Status	0.744	55.536	0.583
	BMI	0.808	9.481	0.246
	S Success	0.069	9.578	0.246
	S SoC	0.075	7.188	0.229
	D Status	0.932	4.322	0.208

Table 7 – Eigenvalues in the Discriminant Analysis

Function	Eigenvalue	% of Variance	Cumulative %	CC
1	2.989	88.3	88.3	0.866
2	0.300	8.9	97.2	0.480
3	0.095	2.8	100.0	0.295

CC: canonical correlation

Table 8 - Classification Results

		Modes	Predicted Group Membership				Total
			PA	Diet	WM	Smoking	
Original	Count	PA	16	6	3	0	25
		Diet	6	13	4	0	23
		WM	1	1	18	1	21
		Smoking	0	0	1	10	11
	%	PA	64.0	24.0	12.0	0.0	100.0
		Diet	26.1	56.5	17.4	0.0	100.0
		WM	4.8	4.8	85.7	4.8	100.0
		Smoking	0.0	0.0	9.1	90.9	100.0
Cross-validated	Count	PA	13	8	4	0	25
		Diet	8	10	5	0	23
		WM	3	3	14	1	21
		Smoking	0	1	2	8	11
	%	PA	52.0	32.0	16.0	0.0	100.0
		Diet	34.8	43.5	21.7	0.0	100.0
		WM	14.3	14.3	66.7	4.8	100.0
		Smoking	0.0	9.1	18.2	72.7	100.0

Table 2 - Cross-correlations of "Physical Activity" Group Variables

	PA A score	PA S score	PA SE	PA Readiness	PA Confidence	PA Risk	PA SoC	PA Success
PA S score	0.291							
PA SE	0.042	0.177						
PA Readiness	0.423	0.362	0.255					
PA Confidence	-0.002	0.254	0.827	0.251				
PA Risk	-0.198	-0.004	-0.078	-0.236	0.060			
PA SoC	0.423	0.362	0.255	1.000	0.251	-0.236		
PA Success	0.129	0.395	0.592	0.636	0.723	-0.034	0.636	
PA Benefit	-0.175	0.140	-0.001	-0.175	0.149	0.850	-0.175	0.074

Table 3 - Cross-correlations of "Diet" Group Variables

	D status	D SE Positive	D SE Negative	D SE Habitual	D SE Temptation	D Readiness	D SE	D Risk	D SoC	D Success
D SE Positive	-0.383									
D SE Negative	-0.405	0.777								
D SE Habitual	-0.331	0.735	0.779							
D SE Temptation	-0.360	0.771	0.621	0.745						
D Readiness	-0.620	0.315	0.350	0.278	0.179					
D SE	-0.354	0.715	0.609	0.680	0.803	0.243				
D Risk	0.241	-0.083	-0.203	-0.187	-0.163	-0.156	-0.120			
D SoC	-0.620	0.315	0.350	0.278	0.179	1.000	0.243	-0.156		
D Success	-0.439	0.461	0.454	0.459	0.431	0.682	0.657	-0.059	0.682	
D Benefit	0.166	-0.019	-0.091	-0.135	-0.073	-0.084	-0.074	0.857	-0.084	-0.051

Table 4 - Cross-correlations of "Weight Management" Group Variables

	Weight	Weight Normal	WM Readiness	WM Confidence	WM Risk	WM SoC	WM Success
Weight Normal	0.424						
WM Readiness	-0.145	0.041					
WM Confidence	-0.290	-0.214	0.029				
WM Risk	0.062	-0.239	0.109	0.060			
WM SoC	-0.145	0.041	1.000	0.029	0.109		
WM Success	-0.137	-0.085	0.683	0.524	0.073	0.683	
WM Benefit	0.077	-0.236	0.033	0.061	0.745	0.033	0.019

Table 5 – Cross-correlations of "Smoking" Group Variables

	S Status	S SE Positive	S SE Negative	S SE Habitual	S Readiness	S Confidence	S Risk	S SoC	S Success
S SE Positive	0.058								
S SE Negative	-0.017	0.959							
S SE Habitual	0.013	0.968	0.961						
S Readiness	-0.566	-0.054	0.012	-0.078					
S Confidence	0.012	0.852	0.824	0.854	-0.316				
S Risk	-0.157	0.493	0.493	0.521	0.097	0.492			
S SoC	-0.150	0.887	0.867	0.868	0.038	0.905	0.552		
S Success	-0.270	0.806	0.796	0.792	0.171	0.788	0.539	0.951	
S Benefit	-0.020	0.397	0.395	0.428	0.011	0.356	0.753	0.345	0.321

Correlation is significant at the 0.05 level (2-tailed).

Correlation is significant at the 0.01 level (2-tailed).

Discussion

Using discriminant analysis an optimal set of independent predictors was identified that determined behavioral choice of users of a computer-mediated decision aid. The resulting set included smoking status (S Status), success estimate in smoking cessation (S Success), smoking self-efficacy (S SoC), body mass index (BMI) and diet status (D Status). Prediction of smoking cessation choice was the most accurate, followed by weight management choice. Physical activity and

diet choices were much better identified in a combined cluster (76%-87%) indicating decisions about these two behaviors were identified by the same variables and the variables that could separate them may have been missing from the dataset. Presence of variables related to individual risks and levels of success in accepting certain health behaviors in the final set of predictors confirmed significance of the computer-mediated decision aid that presented these very variables for user consideration via the TLC DA system. The TLC DA system was instrumental in assisting patients in making informed heath choices. The resulting minimized parameter set can significantly improve user experience by minimizing the amount of information they must manually enter.

Conclusions

Unhealthy behaviors are the most prevalent, preventable and reversible causes of death in the United States [14,15]. They are leading causes of death from cardiovascular disease, cancer and stroke [14]. Physical inactivity and poor diet may have overtaken smoking as the chief preventable mortality risk [12, 13]. Modifiable non-tobacco risk behaviors account for roughly one-third of cancer deaths [16,17]. Additionally, fruit and vegetable intake and physical activity are thought to be independently protective of heart disease [18,19].

The TLC DA differs from previous technologies designed to modify health behaviors. It provides the *opportunity* for choice, rather than an *obligation* to adhere to an assigned regimen. When people perform tasks out of obligation, they subsequently show poorer self-regulation as compared to people who performed comparable tasks that they had chosen themselves [11]. Based on self-determination theory [12, 13], providing a context for making an *informed choice* about which behaviors to change will enhance autonomous motivation and self-efficacy, and result in higher rates of lasting change compared to assigning people specific behaviors to change. Since the average adult practices more than one unhealthy behavior [20], this creates the possibility of intervening comprehensively, efficiently, perhaps even synergistically on multiple risk behaviors simultaneously. To our knowledge, no one has determined predictors of behavior changes people would choose when not constrained. In order to make behavior change research more patient-centered, this is a necessary step.

References

[1] Finkelstein J. et al., Enabling patient-centered care through health information technology, Evid Rep Technol Assess, 2012, 206: 1-531.

[2] Lapshin O, Wasserman E, Finkelstein J. Computer intervention to decrease level of psychiatric stigma among medical students. AMIA Annu Symp Proc, 2006, 998.

[3] Finkelstein J, Feldman J, Safi C, et al. The feasibility and patient acceptance of computer-assisted asthma education in the emergency department, Academic Emergency Medicine, 2003, 10: 505.

[4] Lapshin OV, Sharma K, Finkelstein J. Depression education for primary care patients using a web-based program. AMIA Annu Symp Proc, 2005, 1017.

[5] Skinner C, Finkelstein J. Review of mobile phone use in preventive medicine and disease management. Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies, 2008, 180-189.

[6] Finkelstein J, Wood J, Cha E. Introducing a Blackberry eLearning Platform for Interactive Hypertension Education. Proceedings of the Second International Conference on Mobile, Hybrid, and On-Line Learning (ELML '10), 2010, 77-81.

[7] Finkelstein J, Nambu S, Khare R, et al. CO-ED: A development platform for interactive patient education. Proceedings of the International Conference on Computers in Education, 2002, 648-649.

[8] Kukafka R, Khan SA, Kaufman D, Mark J. An evidence-based decision aid to help patients set priorities for selecting among multiple health behaviors,.AMIA Annu Symp Proc, 2009, 343-347.

[9] Li J, Khan SA, Kukafka R. TLC: an informatics approach to enable patients to initiate tailored lifestyle conversations with providers at the point of care. AMIA Annu Symp Proc, 2009, 364-368.

[10] Chan C, Cohall A, Kukafka R. Selecting data elements to build a patient-centric electronic health record that will support adherence to therapeutic lifestyle change. AMIA Annu Symp Proc, 2008, 101-105.

[11] Heath C, Larrick RP, Wu G. Goals as reference points. Cognitive Psychology. 1999, 38:79-109.

[12] Deci E, Ryan RM. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry. 2000, 11:227-268.

[13] Ryan RM, Deci E. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist. 2000;55:68-78.

[14] Mokdad AH, Marks JS, Stroup DF, et al. Actual causes of death in the United States, 2000. JAMA. 2004, 10;291:1238-45.

[15] McGinnis JM, Foege WH. The immediate vs the important. JAMA. 2004 Mar 10;291(10):1263-4.

[16] Bergström A, Pisani P, Tenet V, et al. Overweight as an avoidable cause of cancer in Europe. International Journal of Cancer. 2001;91:421-30.

[17] Kushi LH, Byers T, Doyle C, et al. American Cancer Society Guidelines on Nutrition and Physical Activity for Cancer Prevention: Reducing the Risk of Cancer With Healthy Food Choices and Physical Activity. CA Cancer J Clin. 2006;56(5):254-81.

[18] Lanas F, Avezum A, Bautista LE, et al. Risk factors for acute myocardial infarction in Latin America: the INTERHEART Latin American study. Circulation. 2007,115(9):1067-74.

[19] Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-52.

[20] Fine LJ, Philogene GS, Gramling R, et al. Prevalence of multiple chronic disease risk factors. 2001 National Health Interview Survey. American journal of preventive medicine. 2004 Aug;27(2 Suppl):18-24.

Address for correspondence

Dr. Rita Kukafka, 622 West 168th Street, Department of Biomedical Informatics, Columbia University, New York, NY 10032. Email: rk326@cumc.columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-113

Towards an Ontology-driven Framework to Enable Development of Personalized mHealth Solutions for Cancer Survivors' Engagement in Healthy Living

Sahiti Myneni, Muhammad Amith, Yimin Geng, Cui Tao

University of Texas School of Biomedical Informatics at Houston, TX, USA

Abstract

Adolescent and Young Adult (AYA) cancer survivors manage an array of health-related issues. Survivorship Care Plans (SCPs) have the potential to empower these young survivors by providing information regarding treatment summary, late-effects of cancer therapies, healthy lifestyle guidance, coping with work-life-health balance, and follow-up care. However, current mHealth infrastructure used to deliver SCPs has been limited in terms of flexibility, engagement, and reusability. The objective of this study is to develop an ontology-driven survivor engagement framework to facilitate rapid development of mobile apps that are targeted, extensible, and engaging. The major components include ontology models, patient engagement features, and behavioral intervention technologies. We apply the proposed framework to characterize individual building blocks ("survivor digilegos"), which form the basis for mHealth tools that address user needs across the cancer care continuum. Results indicate that the framework (a) allows identification of AYA survivorship components, (b) facilitates infusion of engagement elements, and (c) integrates behavior change constructs into the design architecture of survivorship applications. Implications for design of patient-engaging chronic disease management solutions are discussed.

Keywords:

Cancer survivor; Healthy lifestyle; Patient engagement; mHealth; Ontology.

Introduction

The term "cancer survivor" refers to an individual who has been touched by cancer, right from the time of diagnosis, beyond the treatment, through the balance of one's life [1]. The definition also includes family members, friends, and caregivers in addition to the patient, given the way cancer to a loved one can affect their lives. Recent estimates suggest that there are 14.5 million cancer survivors in the United States (US), and the projected numbers indicate a five million increase in the next decade [2]. Worldwide estimates of cancer survivors were 28.7 million in 2008 [3]. However, this reflects only the individuals within five years of diagnosis. In 2006, the Institute of Medicine's landmark report on cancer survivorship highlighted key gaps in the existing care delivery infrastructure and provided several important recommendations for quality survivorship care to this growing population [4]. Cancer survivorship broadly refers to cancer-related physical and psychological issues, lifestyle after treatment such as returning to work, follow-up care, and prevention of secondary malignancies. Survivorship Care Plans (SCPs) provision to every patient is seen as a major step forward in facilitating the care transition from specialists to primary care physicians. These plans can form a communication bridge between physicians and survivors to ensure care quality and continuity, and also act as engagement tools that empower patients to self-manage care.

However, developing a SCP is complex, and a single form cannot address the changing needs of the survivors. The SCPs need to be comprehensive, yet uniquely patient-specific depending on type of primary cancer, age of diagnosis, current age, age-specific late effects of cancer, availability of familial support. In the global context, other aspects including, but not limited to, health beliefs, religious beliefs, and cultural context need to be considered when developing an SCP [5]. Particularly, Adolescent and Young Adult (AYA) cancer survivors' needs demand special attention given the unique circumstances surrounding their survivorship and ever-changing scenarios of life [6]. An AYA survivor's context of survivorship is dynamic given their transition from adolescence to adulthood, physiological and psychological growth, self-identity, separation from parents/family, career pursuits, and involvement in intimate relationships [7]. Therefore, engaging an AYA cancer survivor in health management involves the design of a multi-component solution, where the modules comprising the solution can be used across the care continuum based on patient-specific personalization attributes. Harnessing contemporary technology platforms such as online social media, mHealth applications allow us to generate cancer survivor engagement tools that are cost-effective, ubiquitous, and scalable. In this paper, we provide a novel methodological framework to enable the design of comprehensive and personalized mHealth solutions to engage AYA cancer survivors in healthy living. The framework utilizes ontology to integrate the Patient Engagement Framework (PEF) [8] and Behavioral Intervention Technology (BIT) Model [9] to develop survivor digilegos, which we consider as building blocks of mHealth solutions for AYA survivorship. The unique features of these digilegos are their reusability and customization abilities according to the survivor care context. These digilegos can be put together to build mHealth solutions that are flexible and extensible, yet comprehensive.

Research shows the use of ontology for the design of mHealth applications with a patient-centered focus, specifically for the purpose of self-monitoring [10-12]. Ontologies are suitable for storing and retrieving semantic information, or expressive knowledge, in machine-readable format [13].This information in the ontology can be used to evoke rule-based machine intelligence and decision-making provided through intrinsic reasoning capabilities and querying [14, 15]. Subsequently, PEF acts as a bridging tool to identify opportunities that can help translate the knowledge abstracted within the survivorship ontological representation to engagement elements of a consumer-centered application. Finally, the BIT model allows us to conceive and instantiate the survivor digilegos as technological constituents of the intended

mHealth solution. In the next sections of the paper, we describe the framework components along with illustrative examples demonstrating its application to the domain of AYA cancer survivorship. Specifically, we focus on facilitating the development of mobile apps that engage AYA cancer survivors in follow-up care and healthy living through the digilego components.

Methods

Figure 1 provides a high level representation of our proposed survivor engagement framework. The three main components are (1) Ontology Modeling, (2) Engagement Elements and Opportunities, and (3) Digilegos development using the BIT model.

Ontology Modeling: We constructed an application ontology called "Profile Ontology for Cancer Survivors" (POCS) to model and store knowledge of the patient's after-treatment care plan. The model was sourced from Long-Term Follow-Up (LTFU) guidelines put together for AYA survivors by the Children's Oncology Group. The current draft of the POCS was built on top of the Friend of a Friend (FOAF) ontology model [16], which supported basic profile information of users and agents and also provided a template for associated documents, specifically the cancer survivor plan and its related information. The initial iteration of POCS was authored using Web Ontology Language (OWL), which is a W3C standard ontology language that supports standard knowledge representation and semantic reasoning [17], and was edited with Protégé, which is an open-source, commonly used ontology editor [18].

Engagement Elements: PEF has been developed by Healthcare Information and Management Systems Society (HIMSS) through cumulative layering of five phases- "inform me," "engage me," "empower me," "partner with me," and "support my e-community." A total of nine features have been specified at the highest engagement level, including 'information and way-finding', 'e-tools', 'forms', 'patient-specific education', 'patient access and use', 'patient-generated data', 'interoperable records', 'collaborative care and community support' [8]. This framework was used in our study to sketch the functionality of mHealth infrastructure that potentially facilitates the adoption of SCPs, self-management, goal setting and reinforcement, peer support, and patient-provider communication. A mapping process was conducted to identify the engagement features that could be used to operationalize the POCS ontology classes. The ontological representations derived in the previous step allowed us to understand the characteristics of the information that should be delivered to the user through PEF features. We also specified the level of intended survivor engagement to characterize granularity and complexity of the system features.

Digilegos Development: We used the BIT model to conceptualize digilegos (which are individual and reusable ontology-driven mHealth components) to delineate the operational aims, identify behavior change strategies (where applicable), define user interactions, and outline technical aspects for real-time implementation [9]. The BIT model was originally proposed to develop behavioral interventions, however all digilegos did not have a behavioral component associated with them. For the non-behavioral ones, we still used the BIT model to ensure workflow alignment and smoother interfacing among all the digilegos. Certain digilegos were conceived to be integrated with advanced query and reasoning capabilities of POCS (see Figure 1). In the next sections, we describe the application of the proposed framework to develop AYA survivor digilegos using illustrative examples. Such compartmentalized design architecture of the survivor digilegos allowed us to facilitate customization, harness machine-based querying, integrate engagement tactics, and adopt theory-driven techniques that ultimately enable the development of user-empowering flexible mHealth tools that can be used across the cancer care continuum. Further, the workflow for each of the components has been set toment levels,vors to better mange stress as partof ed the constr

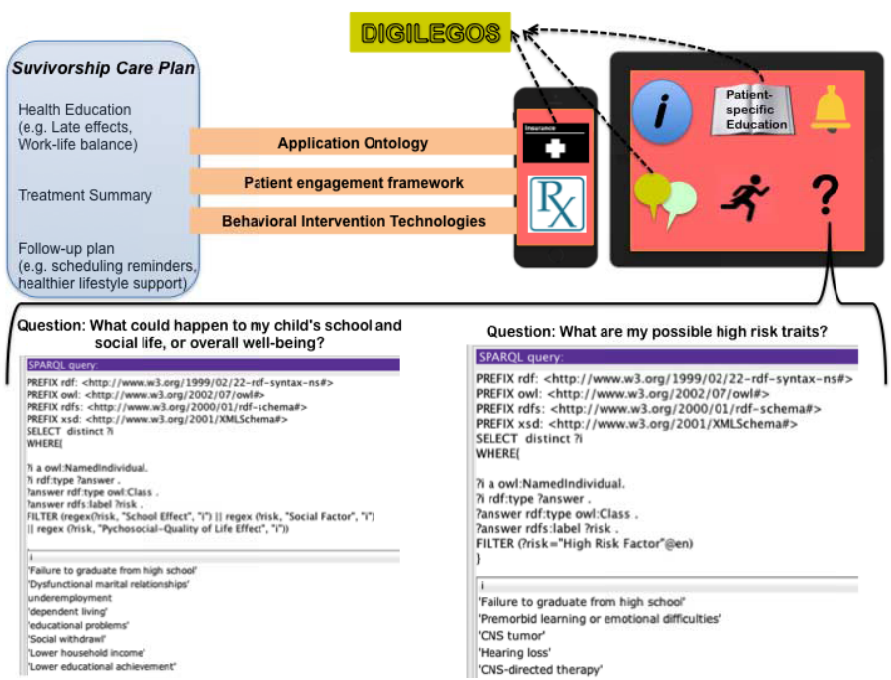


Figure 1- Proposed Survivor Engagement Framework for Development of Personalized mHealth Solution

Results and Discussion

Ontology Modeling: Figure 2 provides a general overview of POCS meta-level ontology. Information abstracted in POCS included end-user's personal information (e.g .marital status, age, gender), the patient's treatment information (complications, treatment type, follow-up schedules, late effects, etc.), and details of personal primary physician and related medical professionals. Initial metrics of our current draft include 58 classes and subclasses, 57 object properties, and 33 data properties. Rule-based intelligence and querying has been used to extract specific answers for survivors' questions posted in lay vocabulary. Examples of questions include "what could happen to my child's school and social life, or overall well-being)?" and "What are my possible high risk traits?" A SPARQL query was used to approximate the question to obtain possible outcomes based on the information represented in the ontology. The integration pipeline for such ontology-driven queries in the proposed mHealth design framework is as follows. Intended steps include: (1) A custom web server application to host the live version of POCS and publish a private interface that will allow our application to query the ontology (2) A separate client-based application ontology that is based on POCS may reside on the application to store personalized information of the user and the user's cancer survivor care information from the hosted POCS knowledge base.

Identification of Engagement Elements: A total of five survivor engagement features were identified using PEF. Given the implementation and domain constraints, the engagement features fell under different phases of PEF. Table 1 provides a list of features and phases that were chosen as part of this step. Most engagement elements identified using the framework are in Level II. This mapping process allowed us to gain insights into the attainable level of survivor engagement, features to be incorporated, and reusable information classes across the features. Few engagement features that have been listed in Table 1 may be pertinent to only certain AYA user groups depending on their gender, age, and cancer type. Examples of such features include the fertility tracker, pregnancy tracker, late effect symptom checker, and lifestyle tips. The PEF framework is intended to guide healthcare organizations in developing and strengthening their patient engagement strategies. Our proposed survivor engagement framework is motivated by the need to support technology development efforts of consumer-centered apps by non-hospital entities. Therefore, certain advanced features listed in the PEF may not be applicable to our study limiting the scope of the engagement elements to Phase II. Such advanced features require constant data sharing with health institutions.

***Digilego* Development:** A total of 11 survivor *digilegos* were conceived based on the engagement features selected using PEF. The individual components cover a variety of survivorship content and engage AYA survivors in care management. For example, "targeted health tips" provide patient-specific education to the survivor taking into consideration their current life scenario from the "personal profile" (e.g. age, gender, marital status). Our POCS ontology model forms the crux for the "late effect summarization" *digilego* to personalize content delivery that is cancer-specific, age-specific, and treatment-specific in nature. Table 2 provides an overview of the *digilegos* and the nature of content provided to the survivor. The overarching operational intention is to promote self-management of cancer care and survivorship among adolescents and young adults. Examples of the sub-goals for each component include increasing positive health behaviors such as physical activity, promoting social well-being through online community, facilitating accessibility to care summary and payment logistics, increasing knowledge

Table 1- Engagement Features and Ontology Relations

Engagement phase	Features	Related ontology classes
Engage (Level II) • Nearest primary care provider • Late effect symptom checker	Information and Way finding	Treatment center; Doctor; Late effects
Engage (Level II) • Fertility trackers • Pregnancy tracker • Social engagement tracker • Healthy eating tracker • Other risky behavioral trackers	e-Tools	Health behavior factor; Patient; Personal Profile; Risk factor; Social Factor
Create synergy and extend reach (Level V) • Care planning (e.g. follow-up scheduling) • Self-management • Daily reminders (e.g. Lifestyle tip)	Patient-specific education	Factor; Late effect; Treatment
Engage (Level II) • Patient profile • Insurance information • Bill pay • Email	Interactive forms	Patient; Personal Profile; Treatment summary
Create synergy and extend reach (Level V) • Online community support forums	Community support	Patient; Social factor; Treatment Center; Doctor

Table 2- Survivor Digilegos and their content specifics

Survivor Digilegos	Content specialty
	Insurance information
	Health behavioral trackers
	Treatment summary
	Personalized late effects summarization
	Follow-up care scheduling
	Personal profile
	Targeted health tips, education
	Transition assistance
	Lifestyle tips; Care reminders
	Social Hub
?	Question corner

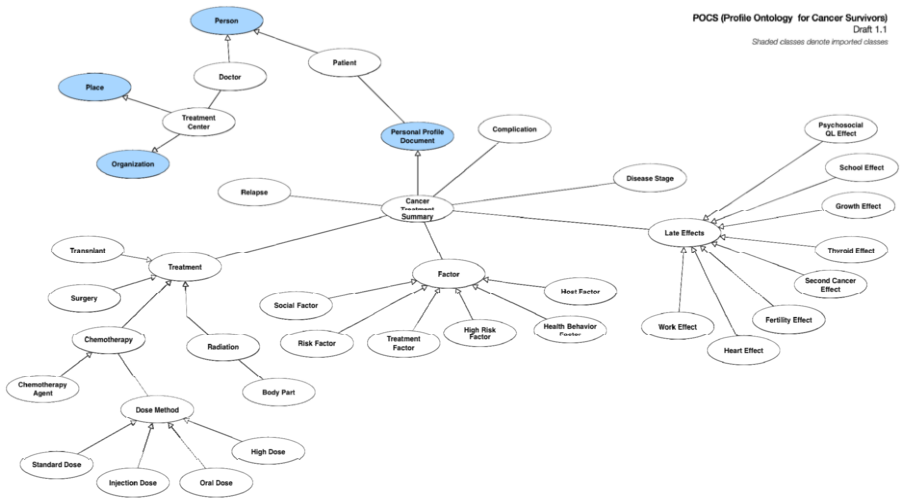


Figure 2- Profile Ontology for Adolescent and Young Adult Cancer Survivors

levels pertaining to late effects specific to certain life events (such as planning for parenthood), and promoting adherence to a follow-up care regimen. The operational intention for each *digilego* is unique depending on the content specialty. The usage intention is the same across all components where the aim is to engage survivors in self-management of their health. Henceforth, social media features, a responsive question corner using ontology-driven methods that can process lay vocabulary, and personalization features have been integrated to the design framework of *digilegos* to promote user engagement.

Several behavior change techniques have been integrated with 7 out of the 11 *digilegos* shown in Table 2. Based on their proposed functionality, the theoretical underpinnings range from education to provide information on health-behavior links, general encouragement through feedback and self-monitoring via e-health tools such as health behavioral trackers, social support through the social hub and community interactions, follow-up prompts to promote self-monitoring behavior, and integration of relapse prevention and stress management tactics for transition assistance and work-life balance [19]. Mapping these strategies to *digilego* elements is straight-forward, given the clear formulation of the intentions of each component and strategy using the BIT model. For instance, "Social Hub" has been mapped to the social support strategy. Similarly, *digilegos* such as "Late Effects", "Patient-specific education" have been designed to provide education on consequences, barriers, and benefits of in self-monitoring and management of cancer survivorship. Further, the instantiation criteria for each of the *digilegos* have been defined in terms of the interaction features and workflow. Notifications, logs, and information delivery are the most used interaction elements. For example, "Work-Life Balance" has interaction features, which include notifications to deliver daily reminders and information delivery mechanisms to provide multimedia educational content to AYA survivors. Similarly, logs have been assigned to "Personal profile", messaging elements and visualizations to "Social Hub" to provide users with (a) communication tools to interact with peer and care providers, and (b) consolidated feedback to users on their social engagement levels. Event-based and time-based workflow criteria have been used to derive personalization effects. For instance, consider a change in education level where the user indicates entering college from high school, event-based criteria are used to deliver appropriate lifestyle tips and transition assistance guidelines that are pertinent to their life in college. Similarly, another event-based workflow is defined for "Late effects" and "Behavioral trackers", where age-specific customizations will come to effect depending on the changes in life scenarios. In summary, a cluster of *digilegos* have been conceived, defined, and characterized using the BIT model to facilitate survivor engagement in cancer survivorship care. The use of ontology has facilitated the identification of abstract survivorship concepts and their interconnections to provide both general and personalized information. PEF and BIT models have provided a way to map and integrate user engagement attributes and theory-driven behavior change strategies.

A feature level comparison of the survivor *digilegos* and existing consumer apps available for AYA survivorship guidance is shown below in Table 3. We considered two apps, AYA Healthy Survivorship [20] and Cancer LateFX [21], developed by Cancer Alliance of Texas and Akron Children's Hospital respectively. Both the apps derive their content from the same LTFU guidelines we used in the study. AYA Healthy Survivorship app focuses on healthy lifestyle promotion while the Cancer LateFX app offers content on late effect management of childhood cancer treatments. Social support features, behavioral goal trackers, profile information and treatment summary management are made available in both the apps and our *digilegos*. While a subset of features delivering patient education, lifestyle tips, care reminders, and a question corner were also found in both the apps, context-specific personalization (e.g. transition assistance) and ontology-driven reasoning capabilities have not been integrated into the existing apps. In summary, the use of detailed ontologies along with the PEF framework and BIT model has allowed rapid construction of survivor *digilegos*, which facilitate behavior change, consumer engagement, and optimal technology use for self-management of survivorship care among AYA survivors.

Table 3- Quick comparison of features in existing consumer apps and the Survivor Digilegos

Apps	Survivor Digilegos
AYA healthy survivorship	
Cancer LateFX	

Limitations and Future Work

The POCS model described in the paper has been created with manual identification of classes. However, a scalable ontology generation will be made possible with the implementation of automated (or semi-automated) natural language processing methods. The use of external sources (e.g. the National Cancer Institute Thesaurus Ontology) allows us to generate comprehensive ontological models. PEF used in the proposed model is aimed at health institutions to develop solutions that are patient engaging. However, the framework is not completely suitable to identify engagement features for technology solutions that solely involve health consumers (e.g. AYA cancer survivors). Future research should focus on consumer engagement taxonomy and a framework that allows technology developers (1) to choose the appropriate level of complexity to model system features, and (2) to identify smoother ontology-to-end user knowledge transfer and query loops. Finally, the *digilegos* described in this paper have not yet been operationalized. Future research should implement and evaluate the *digilegos* in terms of user acceptance, reusability, and technology effectiveness. As part of the next steps, we aim at developing a mobile application comprising the survivor *digilegos* that will interface with POCS using OWL-based API libraries.

Conclusion

Cancer survivorship is a lifelong endeavor for cancer patients, their families, and caregivers. Better survivorship care is vital to enhance the quality of life during and after cancer treatments. With advancements in early detection and treatment protocols, the number of cancer survivors is on the rise around the globe. The proposed survivor engagement framework is a foundational step that will help influence the development of mobile applications with reusable and customizable components as per the needs of the cancer survivors across the care continuum. While the proposed framework is aimed at cancer survivorship, certain *digilegos* can be used to develop mobile solutions for management of any chronic disease (e.g. diabetes, cardiovascular health) that requires patient engagement and self-care. Use of informatics-driven methods, behavior change theories, and patient-centered design approaches can help us develop scalable and cost-effective mHealth products that can be deployed globally.

Acknowledgement

This research is partially supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM011829 to Dr. Cui Tao

References

[1] National Cancer Institute. Office of Cancer Survivorship. http://cancercontrol.cancer.gov/ocs/statistics/definitions.html. Last accessed 12/13/2014.

[2] DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, Alteri R, Robbins AS, Jemal A. Cancer treatment and survivorship statistics. Cancer J Clin. 2014;64(4): 252-71.

[3] Global survivorship statistics. http://www.iarc.fr/index.php. Last accessed 12/13/2014.

[4] Hewitt M, Greenfield S, Stovall E, eds. From cancer patient to cancer survivor: lost in transition. National Academies Press, 2005.

[5] Errico KM, Rowden D. Experiences of Breast Cancer Survivor Advocates and Advocates in Countries with Limited Resources: A Shared Journey in Breast Cancer Advocacy. The breast journal. 2006; 12(s1): S111-S116.

[6] Casillas J, Kahn KL, Doose M, Landier W, Bhatia S, Hernandez J, Zeltzer LK. Transitioning childhood cancer survivors to adult-centered healthcare: insights from parents, adolescent, and young adult survivors. Psycho-Oncology.2010; 19(9), 982-990.

[7] Pratt-Chapman M, Willis A, Bretsch J, Patiemo SR. Improving Cancer Survivorship for Adolescents and Young Adults. Journal of Oncology Navigation & Survivorship. 2013; 4(1).

[8] Patient Engagement Framework. www.himss.org/ResourceLibrary/genResourceDetailPDF.aspx?ItemNumber=28305. Last accessed 12/21/2014.

[9] Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The Behavioral Intervention Technology Model: An Integrated Conceptual and Technological Framework for eHealth and mHealth Interventions. Journal of Medical Internet Research. 2014;16(6):e146.

[10] Villarreal V, Fontecha J, Hervas R, Bravo J. Mobile and ubiquitous architecture for the medical control of chronic diseases through the use of intelligent devices: Using the architecture for patients with diabetes. Future Gener. Comput. Syst. 2014;34:161-175.

[11] Docksteader L, Benlamri R. Mobile Ontology-based Reasoning and Feedback health monitoring system. In: Digital Information Management, 2008. ICDIM 2008. Third International Conference on. IEEE; 2008:876–880.

[12] Docksteader L, Benlamri R. MORF: A mobile health-monitoring platform. IT Prof. 2010;12(3):0018–25.

[13] Kim H-Y, Park H-A, Min YH, Jeon E. Development of an Obesity Management Ontology Based on the Nursing Process for the Mobile-Device Domain. J. Med. Internet Res. 2013;15(7):e130.

[14] Skillen K-L, Chen L, Nugent CD, Donnelly MP, Burns W, Solheim I. Ontological user modelling and semantic rule-based reasoning for personalisation of Help-On-Demand services in pervasive environments. Future Gener. Comput. Syst. 2014;34:97-109.

[15] Bickmore, T. W., Schulman, D., & Sidner, C. L. (2011). A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. Journal of biomedical informatics, 44(2), 183-197.

[16] The FOAF Project. http://www.foaf-project.org/. Last accessed 12/21/2014.

[17] OWL API. http://owlapi.sourceforge.net/. Last accessed 12/21/2014.

[18] Protégé. http://protege.stanford.edu/. Last accessed 12/21/2014.

[19] Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychology. 2008;27(3):379-387.

[20] AYA Healthy Survivorship. https://itunes.apple.com/us/app/aya-healthy-survivorship/id513642187. Last accessed 12/22/2014.

[21] CancerLateFX.https://itunes.apple.com/us/app/cancerlatefx/id725634267?mt=8. Last accessed 12/22/2014.

Address for correspondence

Sahiti Myneni, PhD, MSE

The University of Texas School of Biomedical Informatics at Houston, 7000 Fannin, UCT165, Houston, TX 77030

Sahiti.Myneni@uth.tmc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-118

A Mobile and Intelligent Patient Diary for Chronic Disease Self-Management

William Van Woensel[a], Patrice C. Roy[a], Samina R. Abidi[a,b], Syed SR Abidi[a]

[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Canada*
[b] *Medical Informatics, Faculty of Medicine, Dalhousie University, Halifax, Canada*

Abstract

By involving patients in their own long-term care, patient self-management approaches aim to increase self-sufficiency and reduce healthcare costs. For example, electronic patient diaries enable patients to collect health data autonomously, increasing self-reliance and reducing strain on health professionals. By deploying patient diaries on mobile platforms, health data collection can occur at any time and place, increasing the mobility of chronic patients who typically need to enter health data frequently. Importantly, an opportunity also arises for mobile clinical decision support, where health feedback is directly issued to patients without relying on connectivity or remote servers. Regardless of the specific self-management strategy, patient and healthcare provider adoption are crucial. Tailoring the system towards the particular patient and toward institution-specific clinical pathways is essential to increasing acceptance. In this paper we discuss a mobile patient diary realizing both the opportunities and challenges of mobile deployment.

Keywords:

Self-Management; Mobile Health; Clinical Decision Support Systems; Health Alerts and Notifications.

Introduction

In ageing societies chronic illnesses are becoming increasingly prevalent, negatively affecting people's quality of life and increasing strain on public healthcare systems [1, 2]. To rein in healthcare costs and increase self-sufficiency, patients with chronic illness are increasingly encouraged to self-manage their illness in a home-based setting [3-5]. Related approaches for patient engagement and education involve providing educational and motivational resources [4] and increasing patient self-efficacy by applying behavioural theories [5]. Electronic patient diaries empower patients to autonomously collect health data about their vitals, medications, hospital visits, and symptoms, thus avoiding manual collection by health workers and associated clinic visits. In particular, digital patient diaries accessible through mobile devices (i.e., smart phones and tablets) have the potential to capture patient health data in any setting and at any time. As such, mobile patient diaries provide an up-to-the-minute health profile of the patient and can supply relevant health alerts based on the current profile. Commercial health measurement devices are already accompanied by their own mobile health apps, such as iBGStar and OneTouch Verio Sync blood glucose monitors and iHealth and Withings blood pressure monitors. Moreover, major companies such as Google (Google Fit) and Apple (HealthKit) are developing mobile toolkits for collecting, analyzing, and monitoring personal health data.

As mobile platforms become more sophisticated, new opportunities arise to develop mobile patient diaries that go beyond typical patient data collection and reporting functions. We argue that mobile patient diaries can be extended to incorporate advanced functionalities, offering (i) mobile clinical decision support, (ii) personalized patient interactions, and (iii) data synchronization with centralized electronic medical records. Proactive decision support based on the current patient profile generates relevant interventions to adverse health conditions. By implementing a local Clinical Decision Support System (CDSS), mobile patient diaries are empowered to directly generate timely health alerts, even in situations where wireless connectivity is lacking or the server is unavailable. In previous work [6, 7] we studied how the scale of mobile decision support can be restricted by mobile hardware limitations (i.e., processing, memory). Based on this work we argue that distributed setups are more suitable for mobile systems, in which lightweight, time-sensitive reasoning is deployed locally and heavyweight processes are delegated to the server [2]. For this paper, we studied a lightweight mobile CDSS, configurable with decision rules derived from general or institution-specific clinical guidelines. In doing so, and combined with security and privacy support, we enable healthcare providers to incorporate patient-reported data and mobile CDSS recommendations within their practice.

Connectivity is an important issue in any mobile scenario because lack of Wi-Fi or mobile network coverage may result in temporary or long-term disconnections. One solution is to allow important functionality to be performed offline so disconnections do not limit mobile patient diary usage. As mentioned before, we argue that a mobile CDSS should function independently of connectivity to allow timely health alerts at any time and place. In addition, data entry may be performed offline as well, synchronizing the data with the Electronic Medical Record (EMR) server whenever connectivity is restored. To further ensure the health profile is current and to enable timely health follow-ups, patients should be encouraged to record their health data regularly and respond to alerts raised by the built-in CDSS. We achieve this by supplying personalized and recurring patient interactions, including data entry reminders and health alert notifications. Personalization options include customizing data entry schedules and alert notifications to suit patients' daily lives.

In this paper we present a mobile and intelligent patient diary for managing patients with Atrial Fibrillation (AF). As a core contribution we present a mobile CDSS configured with computerized Clinical Practice Guidelines (CPG) for AF. To improve patient engagement, the patient diary includes a notification engine, customizable with notification schedules that are personalized to suit the patients' lives. We address the connectivity issue by locally deploying a lightweight CDSS, and supporting offline health data collection. Whenever connectivity is restored, data that were collected offline are automatically synchronized with the back-end server. We

implemented the patient diary as a cross-platform mobile application, with current deployments on Android and iOS.

The AF-Mobile Patient Diary (AF-MPD) was developed for the self-management of Atrial Fibrillation (AF) in the context of the Integrated Management Program Advancing Community Treatment of Atrial Fibrillation (IMPACT-AF) project[1]. AF is the most commonly sustained irregular heart rhythm, affecting approximately 350,000 Canadians. Patients carry up to five times the normal risk of developing stroke [8], and stroke associated with AF is almost twice as likely to be fatal [9]. AF-MPD is being used by over 2,000 patients within the Canadian province of Nova Scotia as part of the IMPACT-AF project. Although the patient diary focuses on collecting & acting on AF-related health data in particular, it mainly comprises generic components that are re-usable for the self-management of other chronic diseases.

AF-MPD Architecture Overview

Figure 1 shows an overview of the AF-MPD. Below we summarize the functionality of each component.

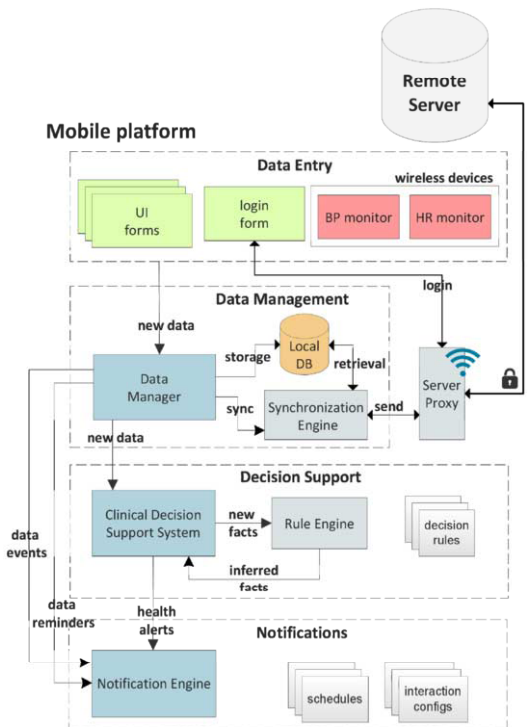


Figure 1 – Architecture Overview

To ensure security and privacy of personal health data, a patient needs to login (**login form**), which requires contacting the server *(login)* for validating the login credentials. Upon authentication, the patient can interact with his AF-MPD by providing health data and reviewing the recommendations. Health data is entered either manually via **UI forms** or automatically via **wireless devices** (e.g., using Bluetooth). This data is passed *(new data)* to the **Data Manager** for persistence in the **Local Database** *(storage)*, and synced with the server *(sync; send)* by the **Synchronization Engine**. If no connectivity is available, the engine will send (*send*) the persistent data *(retrieval)* when connectivity is restored. All server communication occurs via the **Server Proxy**, which maintains a secure server connection. Afterwards, the **Data Manager** sends the new health data *(new data)* to the **CDSS** for inferring new health findings. The **CDSS** incorporates a mobile **Rule Engine** to execute the CPG-based decision rules on the incoming patient data *(new facts)*, and returns the inferred facts *(inferred facts)* as alerts *(health alerts)*.

The **Notification Engine** issues notifications to the patient, including health alerts and reminders. The patient diary components invoke the **Notification Engine** for issuing health alerts *(health alerts)* or registering reminders *(data reminders)*. Such reminders or alerts can be repeated to notify the user frequently until a required action occurs (e.g., acknowledging the alert or performing data entry). In other words, we have implemented persistent notifications that have lifecycles influenced by particular events (e.g., data entry). For this purpose, the **Notification Engine** receives relevant events from other components *(data events)*. As discussed later on, these lifecycles are defined by *schedules* and *interaction configurations*. Below, we elaborate on each component.

Data Entry

Health data may be entered into AF-MPD either manually via UI forms or automatically via wireless measurement devices. A variety of such devices are already available, often accompanied by their own monitoring apps. Open and standards-based Bluetooth communication is available via the Health Device Profile [10], and Bluetooth Low Energy (LE) is typically utilized to conserve battery. However, in practice most devices only allow communication with their accompanying applicationss by applying validation protocols, and manufacturers are not keen on sharing access details (similar issues are mentioned in [3]). Short from hacking their systems, this makes it very difficult to integrate off-the-shelf wireless devices into custom mobile apps. For this reason, the mobile patient diary currently only supports manual data entry.

The AF-MPD supplies UI forms for collecting AF-related health data including heart rate, blood pressure (BP), and relevant symptoms or complications. To enter their current medications, patients can search a local copy of the Health Canada drug and health products database[2]. Additionally, patients can track their costs in dealing with their illness (e.g., clinic visits, loss of work). Graphical overviews of health measurements are also made available, allowing patients to track their progress over time. Figure 2 shows the blood pressure form and the graphical overview of the patient's blood pressure history over time. Figure 3 shows screens for recording AF symptoms.

Data Management

Health data is stored persistently in the local database on the mobile device and synchronized with a remote EMR database server when connectivity permits. Data synchronization may occur in two directions as well. At startup time the Synchronization Engine checks for new patient-specific information in the patient's server EMR (e.g., new medications added by the healthcare provider) and adds them to the local database. To increase patient adherence, the Data Management component registers a persistent reminder to enter health data as well as a synchronization reminder to encourage patients to find connectivity in case data synchronization is overdue.

[1] http://impact-af.ca/

[2] http://webprod5.hc-sc.gc.ca/dpd-bdpp/index-eng.jsp

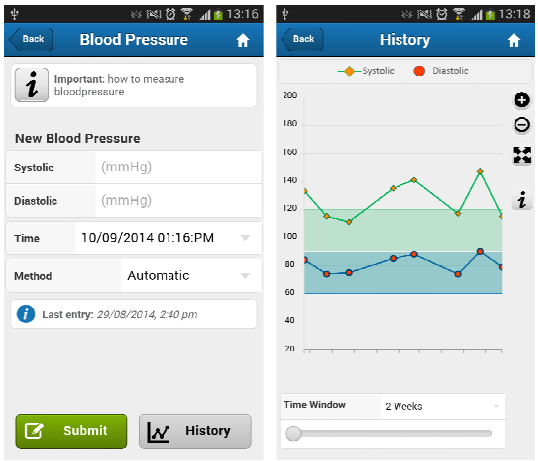


Figure 2 – Blood pressure form and overview

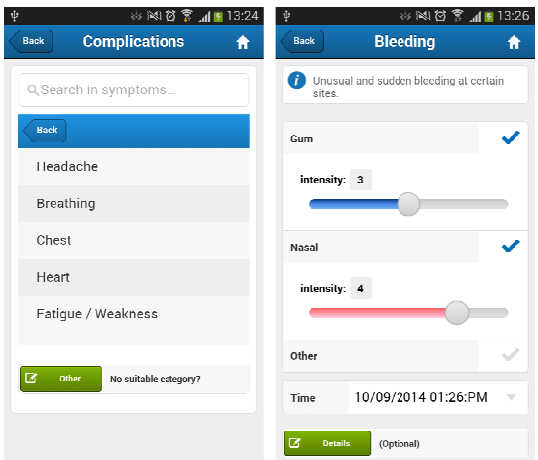


Figure 3 – Symptom browsing/searching and entry form

Server Proxy

Security and privacy are critical concerns when sending health data across the Internet. All server communication occurs via the server proxy, which maintains a secure HTTPS connection with the remote server. On the server side, the valid provenance of health data is ensured via a token system. The server proxy sends the patient's login credentials to the remote server, which returns a unique, temporary token to be included in further communication. Matching the included token to the patient indicated in the health data allows the server to rule out invalid data sources.

Decision Support

By incorporating mobile CDSS, the patient diary can directly issue health alerts based on local health data without requiring connectivity or relying on the remote server. On the other hand, mobile hardware limitations may restrict the scope of mobile decision support. To support large-scale and complex decision logic, the reasoning load was distributed [2], such that we deployed lightweight and urgent processes (e.g., problematic measurements) locally, and delegated more heavyweight processes to the IMPACT-AF server.

In previous work, we developed a framework for benchmarking mobile rule engines [6] and performed benchmarks in the clinical field of AF using off-the-shelf rule engines [7]. Importantly, we observed that in the case of limited reasoning scale and complexity, acceptable performance can already be reached. For the AF patient diary we deploy 11 local decision rules that involve checking for certain symptoms, whether heart rate and blood pressure are within acceptable bounds, and whether dangerous health trends exist (e.g., elevated heart rates for an extended period of time). Since the rules only reference the latest or aggregated health measurements, the reasoning dataset is relatively limited. Patient personalization is supported by parameterizing the rules with patient-specific values (e.g., heart rate limits given the patient's age) that are also included in the dataset. The decision support rules were extracted from CPG for the treatment of Atrial Fibrillation given by the Canadian Cardiovascular Society [11] and the European Society of Cardiology [8]. A key feature of our design is the clear separation between the decision support logic and the application logic; allowing both the rules and application to be updated independently. As such, mobile CDSS can be tailored to suit different, institution-specific clinical pathways and other comorbid chronic diseases.

To illustrate the mobile CDSS, consider a patient who enters high BP measurements over a period of three days with an average systolic BP greater than 135 mmHg and/or diastolic BP greater than 85mmHg (prehypertension). After a measurement is added to the database the rule engine applies its configured rules to infer new health findings. After adding the latest measurement, the relevant rule infers an *uncontrolled blood pressure* finding, indicating prehypertension. In response, the CDSS issues a health alert via the Notification Engine. It can be noted that since reasoning is only applied after adding new health data, the local CDSS has only a minimal impact on battery life.

Notifications

Although health data entry is facilitated by mobile patient diaries, patients still need to frequently perform data entry. In addition, follow-up interventions by physicians are only possible when the collected data is synchronized frequently with the remote server. As a result, a second notification category includes reminders that are issued for data entry and server synchronization. Below we discuss the requirements for these kinds of notifications.

Mobile patient monitoring solutions allow issuing notifications at any time and place. However, this also means patients can be notified in situations where they are not able to perform the required action, such as collecting health data. To account for this constraint in our solution, it is possible to repeat notifications until the desired action can be performed. Since the urgency of the notification will typically increase as time passes (e.g., reflecting an increased necessity for health data entry), we apply a time-delayed strategy that increases notification obtrusiveness over time. This has been achieved through an *event-based* and *evolving* notification lifecycle, where (1) the employed interaction resources (e.g., icons, audio and haptic feedback) increases in obtrusiveness over time, together with notification frequency; and (2) the occurrence of events, such as performing the desired action (e.g., data entry), influences the current state of the lifecycle.

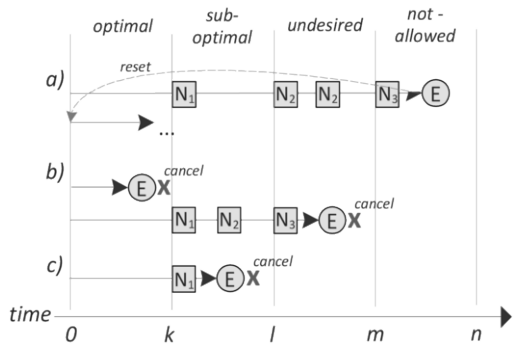


Figure 4 – Example notification lifecycles

In Figure 4 we illustrate several examples of our notification life cycles, including health alerts, data entry, and data synchronization reminders. Notification lifecycles are divided into sequential stages (e.g., optimal, sub-optimal, etc.). To reflect the increased urgency, these progressive stages typically (a) increase the notification's obtrusiveness, by using more intrusive interactions (e.g., sounds and vibrations) and different icons and content templates; and (b) increase the notification frequency per elapsed time. For each progressive stage, interaction configurations and notification frequencies can be predefined for each category (i.e., alerts and reminders). In contrast, registered notifications will each have *individual* lifecycles specifying when each progressive stage starts and finishes; supported events and their effects on the lifecycle; and notification-specific data (e.g., action about which the patient is being reminded) to be filled into content templates.

Each notification lifecycle starts when the notification is registered and is *reset* or *cancelled* when relevant events are received. Figure 4 shows four progressive stages, where the *optimal* stage indicates the initial timespan where no action needs to be taken. The s*ub-optimal, undesired,* and *not-allowed* stages indicate stages with increasing notification obtrusiveness and frequency. In example *a* the system registers a reminder for data entry at install time, where the patient starts receiving reminder instances (N_x symbols) after *k* time has passed. In this case the time *k* depends on the data entry schedule. For example, if the patient needs to enter blood pressures every day, a suitable value for *k* could be 24 hours. Later, the notification frequency and obtrusiveness increase as the *undesired* and *not-allowed* stages are entered until the patient enters the health data. In that case, an event (E circle) is received and the lifecycle is *reset*, starting again from 0. In example *b*, two reminders for data synchronization are registered each time after the patient entered health data while lacking connectivity. In this case, suitable times for *k* depend on the necessity for timely health data uploads (e.g., depending on the urgency of server-side decision support). After *k* time has elapsed, the patient will start receiving reminders of increasing obtrusiveness whereby the start times of subsequent stages (*l* and *m*) will likewise depend on the necessity for timely uploads. For the first reminder, connectivity is re-established before any notification is issued, resulting in an event that *cancels* the reminder's lifecycle. For the second reminder, the patient finally encountered a Wi-Fi hotspot after entering the *undesired* stage, again resulting in the cancellation of the reminder lifecycle. In example *c*, the local CDSS issued an alert after inferring a problematic health issue. In case of alerts, the *optimal* stage will most likely be skipped (setting *k* to 0) so the patient is directly notified of the health issue. In this example, the patient responds immediately after the first notification, generating an event that *cancels* the alert. For instance, patients may respond by simply indicating they will follow-up on the alert, or the alert may include an option to directly dial the clinic or emergency room. The notification lifecycle is terminated after the *not-allowed* stage. We note that notification lifecycles likely depend on the particular clinical setting and chronic illness being managed, and thus should be specified by domain experts.

It is important to note that persistent notifications may result in alert fatigue if not reconciled with the patients' personal schedule [12]. In particular, if patients continuously receive notifications but often are not able to directly perform the associated actions (e.g., health data entry), they will simply start ignoring them. Therefore, we have implemented the functionality to personalize notification lifecycles (within acceptable limits). Currently patients can configure the starting time of the *sub-optimal* stage for *data entry* reminders, thus delaying the time at which they start receiving reminder instances. In doing so, the patient may better align data entry with their own personal schedule. In case this configured time overlaps with subsequent stages (e.g., undesired), the system will display warnings of increased severity during configuration, informing the patient that the current schedule does not allow for an optimal follow-up, but that they should proceed if this is the only timing corresponding to their personal schedule. Setting this time beyond the start time of the *not-allowed* stage is not permitted as it would result in too long a delay for data entry reminders to still have purpose.

Figure 5 shows an example of an Android reminder instance that includes options for directly performing the task ("*Do"*) and personalizing the particular reminder's lifecycle ("*Settings"*). For both our Android and iOS implementations, platform-specific features are utilized to minimize battery usage: on Android, the Notification Engine is only made active by the operating system when notifications need to be issued; on iOS, future notifications to be fired are registered with a platform-specific service. Currently, persistent reminders are registered for data entry and synchronization, while alerts are one-off notifications. Finally, AF-MPD provides patients with a chronological overview of notifications.

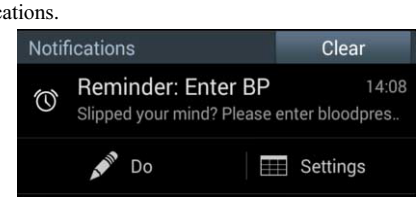


Figure 5 – Example Android notification.

Implementation

The mobile patient diary was implemented using the Adobe PhoneGap [13] cross-platform development framework. In this framework, mobile apps are developed using HTML, CSS, and JavaScript, which are deployed as embedded websites in automatically-compiled native apps. Platform-specific components such as databases, platform-specific notification code, and connectivity listeners are then added to these native apps. Currently, the mobile patient diary is deployed on Android and iOS and the necessary platform-specific components were developed for both platforms.

Related Work

Hommersom et al. [3] examine mobile clinical decision support for multiple chronic illnesses. In order to provide

clinical decision support, they discuss the possibility of using rule-based systems, Bayesian networks, and case-based reasoning. However, this work seems to be in its early stages, summarizing requirements, challenges, and opportunities of mobile smart assistants but not supplying concrete solution designs. Ambroise et al. [2] present a concrete mHealth solution for collecting health data, targeting a group of chronic illnesses. It features a hybrid architecture with lightweight reasoning occurring on the client side and more heavyweight reasoning taking place on the server. Locally collected health data is synchronized at frequent intervals with the server as well. Although the system has an automatic notification system, no details are given on how notifications are managed. At the same time, no consideration is given to increasing patient adherence and the discussed scenarios do not focus on aligning the system with existing, complex clinical pathways.

iALARM [12] is a language for specifying intelligent alerts and introduces the concept of an alert lifecycle. Alerts are triggered by monitoring temporal clinical data whereby alert reminders are fired in case the problem persists and no appropriate alert response has been given. The proposed language focuses on monitoring medical factors in clinics and issuing alerts to healthcare professionals. As a result, reminders only exist in the context of clinical alerts and cannot be issued separately for performing other self-management actions. In the same vein, issues such as notification lifecycle personalization and indicating the increased urgency of notifications (e.g., by increasing interaction obtrusiveness) are not considered.

Conclusion and Future Work

We presented a mobile intelligent cross-platform patient diary for the self-management of chronic illnesses, focusing on AF. Its main features include a local CDSS that allows for timely health alerts without connectivity and a notification system that supports evolving, event-based, and personalizable notification lifecycles. The supplied patient personalization options, combined with CDSS and notification lifecycles customizable towards different clinical settings, allow for improved adoption by healthcare providers and patients. We note that the patient diary comprises generic, configurable components that are re-usable for self-managing other chronic illnesses as well. The mobile patient diary will be used in a large-scale clinical trial (5,000 AF patients).

Future work includes studying additional opportunities for patient personalization, for instance by automatically suggesting suitable data entry times based on analysis of the patient's schedule (e.g., in iCal format). Since only manual data entry is supported at this point, further efforts also involve attempting to plug in wireless measurement devices and using the mobile device's camera to scan medication barcodes and automatically add them to the system. Importantly, we also aim to apply the patient diary to self-manage other chronic illnesses and investigate the extent to which other types of reasoning (e.g., Bayesian networks) would be more suitable for realizing clinical decision support in those cases.

Acknowledgements

This project is funded by a Bayer Healthcare research grant.

References

[1] World Health Organization. Preventing chronic diseases: a vital investment. WHO Global Report, 2005.

[2] Ambroise N, Boussonnie S, and Eckmann A. A smartphone application for chronic disease self-management. In: Proc Conf MobileMed, 2013; pp. 1-4.

[3] Hommersom A, Lucas PJF, Velikova M, Dal G, Bastos J, Rodriguez J, Germs M, and Schwietert H. MoSHCA – My Mobile and Smart Health Care Assistant. In: Proc IEEE 15th Int Conf on e-health networking, applications and services (HealthCom), 2013; pp. 188-192.

[4] Packer TL, Boldy D, Ghahari S, Melling L, Parsons R, and Osborne RH. Self-management programs conducted within a practice setting: Who participates, who benefits and what can be learned? Patient Educ Couns 2012, Apr; 87(1): 93-100.

[5] Abidi SSR, and Abidi SR. An Ontology-Driven Personalization Framework for Designing Theory-Driven Self-Management Interventions. In: Process Support and Knowledge Representation in Health Care, 2013, pp. 97-112.

[6] Van Woensel W, Al Haider N, Ahmad A, and Abidi SSR. A Cross-Platform Benchmark Framework for Mobile Semantic Web Reasoning Engines. In: P. Mika et al. (Eds.) ISWC 2014, Part I, LNCS 8796. Switzerland: Springer, 2014; pp. 389-408.

[7] Van Woensel W, Al Haider N, Roy PC, Ahmad AM, and Abidi SSR. A Comparison of Mobile Rule Engines for Reasoning on Semantic Web Based Health Data. In: Proc of 2014 IEEE/WIC/ACM Int Joint Conf on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). 2014; pp. 126-133.

[8] The task force for the management of atrial fibrillation of the European Society of Cardiology. Guidelines for the management of atrial fibrillation. European Hearth Journal 2010: 31: 2369-2429.

[9] Lin HJ, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, and D'Agostino RB. Stroke severity in atrial fibrillation: The Framingham study. Stroke 1996: 27 (10): 1760-1764.

[10] Health Device Profile (HDP). https://developer .bluetooth.org/TechnologyOverview/Pages/HDP.aspx

[11] CCS Atrial Fibrillation Guidelines Committee. 2014 Focused Update of the Canadian Cardiovascular Society Guidelines for the management of atrial fibrillation. Canadian Journal of Cardiology 2014: 30: 1114-1130.

[12] Klimov D, and Shalar Y. iALARM: An Intelligent Alert Language for Activation, Response, and Monitoring of Medical Alerts. In: D. Riaño et al. Eds. KR4HC 2013/ProHealth 2013, LNAI 8268. Switzerland: Springer, 2013; pp. 128-142.

[13] Adobe PhoneGap. http://phonegap.com/

Address for correspondence

William Van Woensel and Syed Abidi, NICHE Research Group, Faculty of Computer Science, Dalhousie University, E-mail: william.van.woensel@dal.ca & ssrabidi@dal.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-123

Mobile early detection and connected intervention to coproduce better care in severe mental illness

Pauline Whelan [abc], Matthew Machin [abc], Shôn Lewis [abc], Iain Buchan [abc], Caroline Sanders [ab], Eve Applegate [d], Charlotte Stockton [a], Sally Preston [a], Robert Andrew Bowen [a], Zhimin Ze [e], Chris Roberts [a], Linda Davies [a], Til Wykes [e], Nicholas Tarrier [e], Shitij Kapur [e], John Ainsworth [abc]

[a] *The University of Manchester, UK*
[b] *Manchester Academic Health Science Centre, UK*
[c] *Farr Institute @ Health eResearch Centre (HeRC), UK*
[d] *Division of Clinical Psychology, University of Liverpool, UK*
[e] *Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK*

Abstract

Current approaches to the management of severe mental illness have four major limitations: 1) symptom reporting is intermittent and subject to problems with reliability; 2) service users report feelings of disengagement from their care planning; 3) late detection of symptoms delay interventions and increase the risk of relapse; and 4) care systems are held back by the costs of unscheduled hospital admissions that could have been avoided with earlier detection and intervention. The ClinTouch system was developed to close the loop between service users and health professionals. ClinTouch is an end-to-end secure platform, providing a validated mobile assessment technology, a web interface to view symptom data and a clinical algorithm to detect risk of relapse. ClinTouch integrates high-resolution, continuous longitudinal symptom data into mental health care services and presents it in a form that is easy to use for targeting care where it is needed. The architecture and methodology can be easily extended to other clinical domains, where the paradigm of targeted clinical interventions, triggered by the early detection of decline, can improve health outcomes.

Keywords:

Smartphone application; Mobile Health; Mental health; Self-management; Connected health.

Introduction

Severe mental illness, including what is diagnosed as schizophrenia and psychosis, usually appears early in adult life and has a high risk of relapse. Relapses often result in unscheduled hospital admissions, with substantial suffering of the individuals affected and their families, and high costs for mental health services. Current approaches to the management of mental illness are limited by sporadic monitoring that takes place during consultations with mental health staff who ask service users to recall symptoms from as much as a month earlier. This has led to problems with averaging, reliability and recall, and has made accurate clinical assessment difficult [1]. Maintaining remissions and supporting recovery represents a major challenge for mental health services, which have lacked an approach that is personalised, responsive and service user-focused. Moreover, a recent report by the Care Quality Commission, the independent regulator of health and social care services in England, noted how service users often do not feel involved in their care planning [2], while National Guidelines stress the importance of service user self-management of symptoms [3]. Over recent years there has been an explosion of mobile, ubiquitous and connected health technologies, which can be used to detect diagnostic signals from individuals with much greater precision and resolution than ever before. These technologies promise to transform healthcare [4] and could support the coproduction of care between service users and care professionals [5].

In severe mental illness, the early detection of a change in symptoms can be used to trigger early intervention and so prevent relapse [6,7]. We have extended the standalone ClinTouch app [1] to close the loop between service users and care teams to enable the early detection of relapse and the targeting of interventions. We anticipate that this will improve patient outcomes by preventing relapse and hospitalisation. A reduction in hospitalisation will result in costs savings for healthcare providers.

Methods

To our knowledge, ClinTouch is the only system in the UK that offers a validated mobile assessment technology for severe mental illness with real time symptom reporting, online monitoring, relapse detection and the proven ability to integrate with existing mental health services. The use of smartphone apps within mainstream mental health care services in the UK is nascent and innovative and current clinical approaches to the management of mental illness typically rely on infrequent face-to-face consultations. The clinical aims of the ClinTouch system are to provide improved modes of personalised clinical monitoring and to enhance service user self-management. From a technical standpoint, the challenges include: a) developing a secure system that can collect and monitor service user symptoms in real-time; b) providing a reliable system to store service user clinical data; c) allowing flexible integration with diverse electronic clinical record systems across providers in the UK National Health Service (NHS); d) providing a way for clinical staff to easily view and monitor service user data; e) presenting summaries and visualisations of data that are relevant, understandable at a glance, and actionable; f) detecting deterioration in symptoms and risk of relapse; and g) supporting clinical research by making collected data available to researchers.

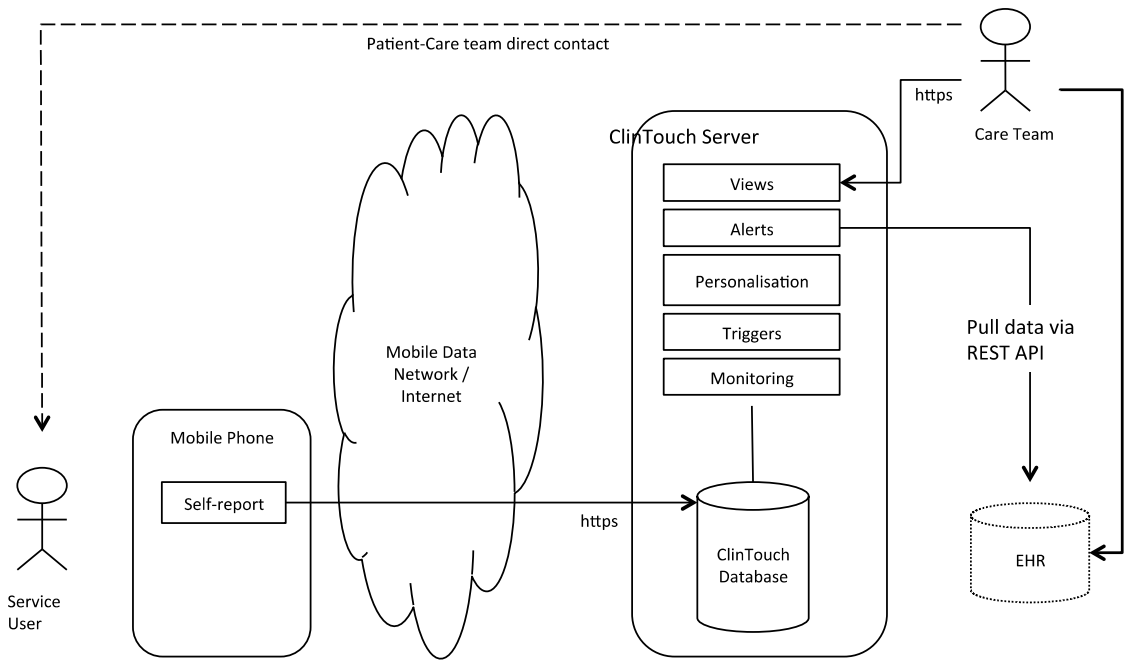


Figure 1 – Overview of ClinTouch system architecture. Service users use the app on their smartphone to capture continuous, longitudinal symptom data, which is transmitted in real-time through the mobile phone network to the central ClinTouch database for persistent storage, hosted on a backend server. This server also hosts a process running the clinical monitoring algorithm and provides a web-based interface for health professionals to configure service user accounts and to view data and alerts. Another process on the server manages data exchange between the ClinTouch system and the electronic health record system.

The architecture of the ClinTouch system is shown in Figure 1. The core elements of the system include: a smartphone application which allows service users to record their symptoms several times a day and wirelessly uploads the data to a central secure ClinTouch server; a web-based interface, which allows mental health staff within the provider organisation to view the uploaded symptom data, and a custom clinical alert algorithm, which is run over the collected data. All communication to and from the ClinTouch server (from the app, from the web-based client and via the REST API) is secured over https. All symptom data are stored against a unique participant identifier on the phone, uploaded as soon as possible and then stored securely in the ClinTouch database.

An agile, user-centred design methodology was adopted throughout the project. An iterative approach to development was supported by a series of focus groups conducted with care staff and service users, who provided input on requirements and qualitative feedback on software prototypes and designs.

Smartphone Application

A smartphone application prompts service users to record their symptoms several times a day at pseudo-random intervals. Data are then wirelessly and automatically uploaded to the ClinTouch server. The mobile assessment technology uses gold-standard rating scales and has been shown to be reliable and valid for assessing psychotic phenomena [1]. Drawing on Sama et al's [8] categorisation of user engagement, derived from the transtheoretical model of change and behaviour modification models, the app supports self-monitoring and progress tracking, which are achieved through collection and summary displays of symptom data.

Figure 2 – A screenshot of the question response page on the smartphone application.

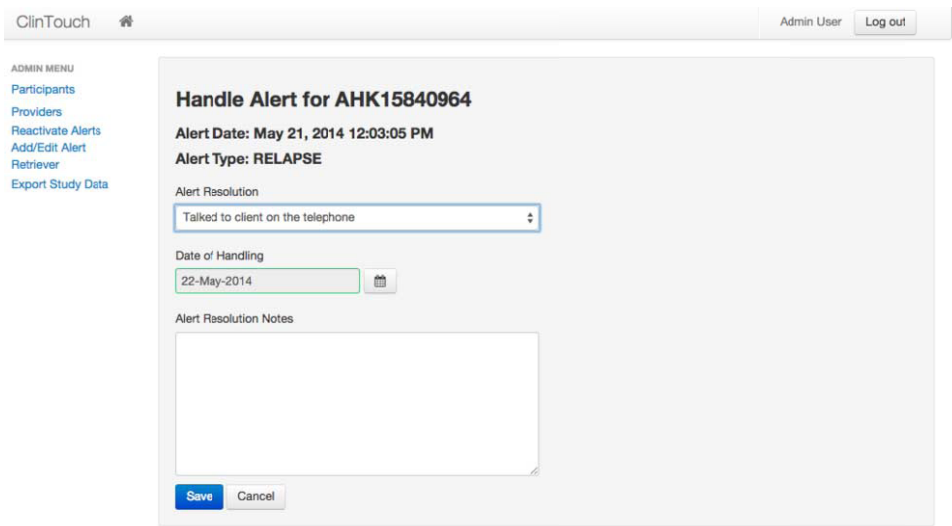


Figure 3 – A screenshot of the ClinTouch web interface: web page shows the screen for handling an alert raised by the clincial algorithm when it detects a high risk of relapse.

Web Interface

A web-based interface allows mental health staff to securely view the symptoms recorded by the service users in their care. The web interface supports multiple user roles, restricting access to service user data based on the logged-in user and allowing researchers to export symptom data for further analysis. The web interface provides configuration screens for adding and editing service user and staff accounts to the ClinTouch system. When adding service user accounts, questionnaires to be displayed on the smartphones for symptom monitoring can be personalised.

Clinical Algorithm

A custom-developed clinical algorithm is run across the symptom data to detect risk of relapse. If a high risk of relapse is detected, an alert is raised in the system. ClinTouch can be configured to either send an alert email automatically or third parties can integrate with the system through a REST API [9] provided by the ClinTouch server. In addition to email notification, the alert status of a service user (ALERT or OK) is viewable via the web interface. A raised alert must be manually cleared by the care provider via the web interface, where an alert resolution must be entered for auditing purposes (see Figure 3). The alert algorithm is highly personalised for individual service users at the time of configuration of the service user within the ClinTouch system. The weighting for symptoms for relapse used in the algorithm can be further edited after the participant has started using the smartphone application. This allows for ongoing personalisation following clinical assessment.

Results

The ClinTouch system has been developed based on findings from two pilot studies which used the smartphone app. The first pilot study confirmed that symptom reporting within the smartphone app is a valid way for assessing psychotic phenomena [1]. This pilot study involved 44 participants and used six pseudo-random alarms per day, for one week. Compliance (defined as completion of at least 33% of datapoints over seven days) was found to be 82% among participants. The second pilot study extended the time period for which participants used the smartphone app to six weeks and involved 7 participants. Of the 6 participants who completed the study, compliance was 80%. The majority of participants reported finding the app helpful.

The current ClinTouch system embeds the smartphone app within the loop of care to enable real-time clinical monitoring and detection of relapse. The system has been successfully deployed in two English NHS provider organisations. At the time of writing, 34 service users and 12 staff members are actively using the system as part of a randomised controlled trial to assess the system's feasibility for use in clinical practice. The total number of participants involved is set to rise to 40 during the course of the trial.

Input from staff and service user groups throughout the development of the software has ensured the software is meeting the needs of end users. Feedback from our user groups highlighted the importance of an engaging user interface for the smartphone application. Development resources were consequently invested in this area, leading to a professionalised smartphone application and ensuring quality of the user experience. Service user focus groups led to increased personalisation of the smartphone application, including a) the ability to enter contextual data to accompany symptom data and b) the ability to customise the look and feel of the application. The design of the web interface was guided by staff input, who defined the formats used in the display of symptom, alert and export data.

As NHS provider organisations in England do not have a standardised technical platform for maintaining electronic clinical records, the software was developed with the flexibility to integrate both with diverse electronic clinical record systems and to operate as a standalone solution. To support integration, a secure REST API [9] allows secure pulling of the alert data from the ClinTouch system by a third party. The system can also operate in standalone mode with the web-based interface operating as a fully functional platform with a comprehensive view of service user data. In practice, both solutions proved useful: one NHS provider organisation integrated ClinTouch with their electronic clinical record system prior to the start of the trial; at the second site, delays in the development of the local electronic

clinical record system meant that the ClinTouch system has been used in standalone mode.

While qualitative feedback from staff who have engaged with the system has thus far been largely positive, strains on mental health services generally in the UK have meant that it has been difficult for some staff to find the time to use the system. It is hoped that as more staff engage with the system, the value of ClinTouch for enhancing care will be more widely recognised and will drive further engagement. Service user feedback highlighted the need for further personalisation of the app and future development will focus on this area.

Discussion

The ClinTouch system was designed to address the gaps in current care provision. The problem of intermittent symptom reporting is overcome in ClinTouch by incorporating a validated assessment tool within the smartphone app where symptoms can be recorded in real-time. Service users can use the app to self-manage and monitor their symptoms and work collaboratively with their care providers which could help address previously-reported feelings of disengagement. As symptoms are monitored in real-time by a clinical algorithm which alerts care staff when risk of relapse is high, ClinTouch could facilitate earlier intervention and reduce hospital admissions.

For clinical staff, ClinTouch provides enhanced, personalised clinical monitoring. For service users, ClinTouch provides tools to assist with self-management. For mental health services, the system can reduce costs by facilitating early intervention, and reducing unscheduled hospital admissions. ClinTouch has been successfully deployed across two English NHS provider organisations and aims to improve the care of individuals who experience psychosis.

ClinTouch has proven ability to both integrate with existing electronic health records systems and to operate as a standalone platform. The platform provides a rich feature set, including: display and capture of questionnaire data via a personalised smartphone application; wireless uploading of questionnaire response data; a web-based interface for viewing symptom data in tabular and dynamic graphical format; a data export facility for researchers; and a mechanism for generating alerts based on a personalised clinical algorithm for detecting risk of relapse. Ongoing challenges include engaging a broader care provider population and providing further personalised modules within the app.

Conclusion

The ClinTouch system was developed to close the loop between service users and health professionals, and to address key limitations of current approaches to the management of severe mental illness. ClinTouch provides an end-to-end secure platform, incorporating a validated mobile assessment technology, which allows service users to record symptoms in real-time; a web interface to display service user data to service users and care staff; a clinical algorithm that runs on submitted symptom data to detect risk of relapse; and an alerting system to flag the need for clinical intervention. ClinTouch integrates high resolution, continuous longitudinal symptom data into mental health care services and presents it in an intuitive format that enables clinical staff to target care where it is needed. The system is currently being trialled in two NHS provider organisations in England. The architecture and methodology can be easily extended to other clinical domains, to enable early detection of decline and improve health outcomes.

The ability to identify the need for early intervention through real-time personalised monitoring has widespread applicability across a range of health conditions, where early intervention is key to preventing adverse outcomes. The ClinTouch platform offers a psychometrically well-founded customisable framework for real-time symptom capture, data visualization and analysis, which has been successfully deployed in mental health and is ready to translate to other contexts of co-produced self and professional care.

References

[1] Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, Rogers A, Wykes T, Kapur S, Buchan I, Salter E, Lewis SW. The feasibility and validity of ambulatory self-report of psychotic symptoms using a software smartphone application, *BMC Psychiatry*, 2012; 12:172. doi:10.1186/1471-244X-12-172.

[2] Care Quality Commission, Monitoring the Mental Health Act in 2012/2013, 2014. http://www. cqc. org. uk/sites/default/files/documents/cqc_mentalhealth_2012_13_07_update. pdf (accessed 18 December, 2014).

[3] National Institute for Health and Care Excellence, Psychosis and Schizophrenia in adults: treatment and management, 2014, http://www.nice.org.uk/guidance/cg178/resources/guidance-psychosis-and-schizophrenia-in-adults-treatment-and-management-pdf (accessed 18 December, 2014).

[4] Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res, 2013, 15(5):e:95. doi:10.2196/jmir.2430

[5] Boye N. Co-production of Health enabled by next generation personal health systems. Stud Health Technol Inform, 2012, 177:52–8.

[6] Birchwood M, Spencer E, McGovern D. Schizophrenia: Early warning signs. Advances in Psychiatric Treatment. 2000, 6(2):93-101. doi: 10.1192/apt.6.2.93.

[7] Gleeson JF, Rawlings D, Jackson HJ, McGorry PD. Early warning signs of relapse following a first episode of psychosis. Schizophr Res, 2005, 80(1):107-11.

[8] Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An Evaluation of Mobile Health Application Tools. Eysenbach G, ed. JMIR mHealth and uHealth. 2014;2(2):e19. doi:10.2196/mhealth.3088.

[9] Fielding RF. Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000. http://www. ics. uci. edu/~fielding/pubs/dissertation/rest_arch_style. htm (accessed 18 December, 2014).

Address for correspondence

Dr Pauline Whelan
Centre for Health Informatics
Institute of Population Health
Vaughan House
University of Manchester
M13 9PL
England
pauline.whelan@manchester.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-127

Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability?

Mobin Yasini[a], Guillaume Marchand[a,b]

[a] *DMD santé, Research and development department, 51100, Reims, France*
[b] *Department of psychiatry, Rouen University Hospital, 76031 Rouen Cedex, France*

Abstract

Health-related mobile applications (apps) have been shown to improve the quality of health and patient care. Their use in clinical and health-related environments is becoming more considerable. The number of health-related apps available for download has considerably increased, while the regulatory position of this new industry is not well known. Despite this lack of regulation, measuring the usability score of these apps is not difficult. We compared two samples of twenty health-related applications each. One of the samples contained the apps with top-rated usability scores, and the other contained the apps with lowest-rated usability scores. We found that a good usability score correlates with a better medical reliability of the app's content ($p<0.005$). In the period in which a valid regulation is still lacking, calculation and attribution of usability scores to mobile applications could be used to identify apps with better medical quality. However, the usability score method ought to be rigorous and should not be rounded off with a simple five stars rating (as is the case in the classic app stores).

Keywords:

mHealth, Mobile Applications, Regulation, Smart Phones, Equipment Safety.

Introduction

Mobile technology solutions have an important place in daily life today. In addition to simple communication and text messaging, a variety of applications running on smartphones and tablets could offer easily accessible solutions to almost any possible requirement that may arise in the field of information and communication. Health-related information and communication is not an exception in the exploding penetration of mobile technology phenomenon.

Mobile Health – or mHealth – is the practice of medical or public health supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [1]. Service care providers, researchers, and national governments are excited at the opportunities mobile health has to offer in terms of improving access to health care, engagement and delivery, and health outcome [1]. Numerous functionalities including data collection [2], patient communication [3], health care delivery [4,5], and patient and professional education [6,7] have been used to support patients or healthcare professionals.

Currently, more than 97,000 mHealth applications are present in the classic app stores, and approximately 1,000 new applications are being published every month [8]. In the next few years, more than 3,000,000 free downloads and 300,000 paid downloads are expected to be made of mHealth applications just in the USA [9]. Being easily accessible and highly available make smart devices (smartphones and tablets) very attractive for consumer and professional areas of application [8]. The smartphone is considered to be the most popular technology among physicians since the stethoscope [9]. Therefore, there are high expectations for growth in the mHealth market. With this growing market and the overwhelming diversity of applications, there is a need to recognize which applications are safe and trustworthy. In the recent literature, potential dangers and reliability of some health-related applications were investigated [10–12]. Various aspects of mobile health applications have been called into question, including medical professional involvement in the development of the applications [13], the accuracy and reliability of the content of the applications used in diagnosis and patient management [14], the potential danger of camera function in mobile devices to judge weather skin lesions are suspicious [15,16], deficiencies in self-management applications in diabetes [17] and asthma [18].

The mobile health application industry is still in its infancy. Existing laws and public regulations for approving health-related applications are only relevant to a rather small number of applications [8]. Happtique, a private certification of health-related applications suspended its operation after a developer discovered that two certified apps contained data insecurity issues [19]. Various reasons, including the number of functionalities, diversity of information and rapid development of health-related applications in the market, make certification difficult to achieve [20].

A usability score indicates the satisfaction and effectiveness of an application. Major platforms for mobile applications (Google Play store & Apple store) provide consumers the ability to rate the applications. Users can rate the applications by giving them a mark out of five stars. In the absence of a comprehensive and rigorous certification of mobile applications for health care, the only easily accessible quantitative measurement is the usability score. The main aim of this study was to evaluate if good usability scores correlates with a better medical safety. We therefore set out to determine if health-related apps with disparate usability scores differ in medical safety aspect.

Methods

French language-based health-related mobile apps in classic app stores (Google play store and Apple iOS app store) were gathered on our website at www.dmdpost.com. A monitoring system based on keywords extracted health-related applications from the classic app stores. App editors who would like to have their applications evaluated by our community could also directly send us their apps. The website currently contains 896 mobile health applications. These applications were divided into two categories according to

their target user: health professionals and the general public. A panel of application users (health professionals and the general public including patients) voluntarily completed usability satisfaction questionnaires for each mobile application. Everyone (health professionals, patients, and healthy individuals) could evaluate the apps addressed to the general public. However, if the target users of an app are health professionals, it would be evaluated solely by its target users. There were three questionnaires according to the evaluator's profile:

- A questionnaire for patients or healthy individuals to evaluate the apps intended for the general public
- A questionnaire for health professionals to evaluate the apps intended for the general public and to identify whether the health professional is willing to prescribe the app to his/her patients
- A questionnaire for health professionals to evaluate the apps intended for health professionals to verify if the application matches the healthcare professionals needs

There are two types of questions in each satisfaction survey both in a closed-ended format:

- Influence questions: to quantify application usage and reliability of answers (for example, Did you use the application several times a day for more than one week?)
- Perception questions: to assess the quality of the application on its functionality and ergonomic aspects (for example, Do you find the application easy to use?)

When the number of evaluations reaches five (at least four individuals and one health professional), the embedded algorithm assigns a score out of 20, judging the overall quality of the application. If an application is updated, it will pass to the status "Rating in progress" and will be re-evaluated by at least four individuals and one health professional to mark its current version.

We then selected the 20 highest-rated and the 20 lowest-rated applications. These applications constitute the material for this study. Two medical doctors working on mHealth evaluated four basic issues to assess medical safety of these applications. The four issues are:

- Prescence of at least one related health professional in the conception or development of the application
- Using reliable and valid bibliographic references to create the application contents
- Clear identification of objectives, target audience, and/or mentioning the possibility (or no possibility) of substituting a medical advice within the application
- Validation of the application in a peer-reviewed scientific paper

These criteria were then validated by a group of academic experts in the relevent field. The best rated and lowest rated applications were listed randomly so that the evaluators did not know the applications' usability scores during their evaluation process. The evaluators examined independently each application with the basic criteria. Discordances were then discussed to obtain a consensus. The name of each application was searched in Medline and Web of Science databases to identify if the application was the subject of a peer-reviewed published research. We then compared the total scores in the two groups. Descriptive information was produced for each basic issue of medical safety. The results were analyzed using Chi-square test. Significance of the statistic test was set at $p<0.05$.

Next, we compared the given app store usability scores of the two groups. In this step, we were interested in finding out if the user evaluations out of five stars could be considered as a reliable criterion to predict the medical reliability of the applications. The user evaluations out of five stars, published in the classic app store platforms within the app description, were analysed and compared in the two groups. We used the Wilcoxon rank-sum test to compare the mean score in the two groups.

Results

The applications used in this study covered various subjects including allergies, contraception, pregnancy and menstrual cycle, calorie intake, bipolar disorder, multiple sclerosis, medical or therapeutic education, cardiology scores, general follow up, diabetes mellitus, traveler's health, and hepato-gastroentrology recommendations. The target users of these applications were health professionals and the general public including patients. The majority of apps in both groups were intended for the general public. However, the number of apps intended for health professionals was more prevalent in the highest-rated group than the lowest-rated group (40% of the apps in the top-rated group against 5% of the apps in the lowest-rated group).

The number of concordances with our medical safety criteria together with the relevant percentages are illustrated in table 1.

Table 1– Prevalence of fulfilled medical trustworthy criteria in top rated and lowest rated applications

	Highest rated applications (%)	**Lowest rated applications (%)**
Presence of related health professional	14 (70%)	8 (40%)
Reliable references	11 (55%)	0 (0%)
Identification of objectives, target audience, possibility of replacing a medical advice	16 (80%)	16 (80%)
Validation of the application in a peer reviewed scientific paper	0 (0%)	0 (0%)
Total	41 (68.3%)	24 (40%)

Prescence of related health professionals in the conception or development of the application was observed more in the top-rated applications than the lowest-rated applications. The criterion satisfied in the fewest number of applications in both highest-rated and lowest-rated groups was mentioning whether reliable or valid bibliographic references were used to create the application contents (11 in top-rated applications and 0 in lowest-rated applications). The third criteria fulfillment (idenitification of objectives) was observed equally in both top-rated and lowest-rated applications (16 out of 20 for each group). In the top-rated group, we found a total of 41 times (68.3%) satisfaction to our medical safety criteria versus 24 times (40%) in lowest-rated applications. None of the studied applications were the subject of a peer-reviewed paper to assess its effectiveness. We therefore decided to remove this criterion from our calculations.

According to our criteria, the medical reliability was significantly higher in the highest-rated applications compared to the lowest-rated applications ($p<0.005$).

In comparing an application's usability score to the classic app store rating, we found that the highest-rating group have a mean score of 3.87 (standard deviation = 0.67). In the highest-rated group, four applications were removed due to an insufficient number user ratings for the application. The range of marks was between 3.5 and 5. In the lowest-rated group, six applications were removed due to insufficient user ratings from the classic app store. The mean score for the 14 remaining was 3.78 (standard deviation= 1.03). The range of ratings in this group was between 2 and 5. We did not find a statistically significant difference between these two means ratings ($p<0.9$) in our samples. We therefore did not compare the store provided usability scores with our medical criteria.

Discussion

Mobile applications comprise a flourishing industry today. Health-related applications provide users with an unprecedented opportunity to achieve a better quality of health as well as a better quality of healthcare. However, when it concerns health and healthcare goals, one must consider the potential risks and dangers. The ubiquity of mobile applications – including health-related apps – brings additional risks to less-experienced users who might download applications with dubious medical information and advice [21]. Although some voluntary app certification schemes like "Happtique Health App Certification Program" exist [21], a worldwide comprehensive and rigorous certification or regulation in all possible aspects concerning the mobile health applications including privacy, operability, security, ethical issues, medical content reliability, etc., is still lacking. The traditional methods of evaluation are not suitable for the fast paced nature of technology [22]. However, calculating a usability score is more common and has already been used in the published investigations [17,23–26]. In this study we present an analysis of 40 health-related applications to assess the feasibility of judging the medical reliability of mobile health-related applications by their usability score. In particular, we developed a usability evaluation method that was more rigorous than the present rating system out of five stars available on the classic store platforms. We defined four basic criteria to assess the medical content of health-related applications. We found that these criteria are significantly more fulfilled in the applications with a better usability score than the applications with lower usability score.

Other studies that evaluated the usability of health-related applications admitted that good usability scores in applications are related with better results in functionality and high compliance of consumers [17,23–26]. This corroborates the results of our study, which report that good usability scores correlate with better medical safety.

We then compared the store platform provided usability score in our samples. Our findings showed that the store-provided usability score was not significantly different in the two groups. A simple consumer rating out of five star is therefore not a good benchmark for evaluation of the usability of health-related apps.

This study had several limitations, which we tried to mitigate. We blinded evaluators of applications' usability scores to reduce bias during their evaluation of medical criteria.

The rate of being approved by a peer-reviewed publication was poor in both highest-rated and lowest-rated groups of apps. This is in line with another published study in which among 283 pain-related apps, none of them had been scientifically validated or proven to be effective [9]. However, our sample applications were only in French language. This can reduce the chance of being the subject of a peer-reviewed publication, because most of these publications are in English language and assess rather the English language applications.

The use of reliable references in the app development was not sufficiently mentioned for any of the apps, even in the highest-rated applications. This defective feature could be somewhat compensated by having a licensed health care professional implicated in the development of the application. Furthermore, valid and reliable sources might have been used in some applications without being mentioned.

Prescence of related health professionals in the development of the application was found to be the most prevalent source of support to health-related apps. This finding corroborates with another study who reports the implication of health care professionals in the app development followed by being recommended by a patient association as the most important source of support to the apps [9]. An analysis of colorectal disease smartphone apps in another study showed that only 29% of the apps had had customer satisfaction ratings and 32% had named medical professional involvement in their development or content [13]. This finding is in line with our results in lowest-rated health-related apps with 40% of medical professional involvement.

The present user evaluation system provided by the store platforms seems not to be a good reference to conclude if a health-related app is medically safe. This study demonstrated that a more rigorous system to measure usability score of the health-related apps could correlate with a better medical reliability. It is not yet clear how the assessment for health-related apps should best proceed [22]. In this situation, although the correlation with usability score is not an absolute and precise solution, it can be used in the absence of a comprehensive regulation.

The proportion of apps intended for health professionals was found to be more prevalent in the highest-rated group. It seems that app editors or developers pay more attention to the content when the target consumers are health-literate professionals. However the majority of health-related applications are intended for the general public (as is also evident in our samples) who may not have the same health literacy and numeracy of health professionals. Educational programs for health-related app users could not only empower patients and healthy individuals to have a better insight of the apps that they are using, but also arouse app editors to develop better quality and reliable apps.

Perhaps the most important limitation of this study is that we only looked at French language applications. This could have limited the yield of the results. However, some of these apps were present in the stores in a multilingual format. Small sample size is another limitation. We selected only 40 of the possibilities, not only because it was convenient, but also because it was what could be feasibly done. In this study the apps were taken from all the diverse categories of health related apps that could be found. As the different health fields have different characteristics, further studies analysing apps by dividing them according to their purposes or functionalities would be of considerable interest.

Even in the top-rated group, a great percentage of applications (more than 20%) lacked at least one of the medical criteria. This implies the need for a comprehensive multi axial program to assess the health-related apps in all aspects. Medical trustworthy of health-related apps is of utmost importance, however high-quality, evidence-based content

alone is of limited value, if presented in a way that does not adequately match and address the usability, accessibility, readability (reading with understanding) and health literacy needs of targeted audiences [21].

Conclusion

Mobile health technology seems to be a logical, acceptable, and efficient way to improve the quality of health and health care services. It can make healthcare more accessible and affordable for all [9]. It has been demonstrated that smartphone-based applications improve the efficiency of healthcare delivery and make healthcare more effective [27]. To achieve this objective, it is clear that the content of applications used in a medical or health context are required to be trustworthy. Mobile health applications still comprise a new industry. Therefore, it takes some time to achieve a valid comprehensive assessment procedure to proceed the app review. Meanwhile, calculation of usability scores that are more common among mobile applications could be used to distinguish apps with better medical quality. However, the usability score method ought to be rigorous and should not be rounded off with a simple five stars rating. Further research with more important samples would help to confirm the results of this study and pave the way to find trustable methodological issues to assure the quality of mobile health.

Acknowledgments

The authors gratefully acknowledge the contribution of Mr. Gregory Duval and Mr. Bastien Burlion for their recommendations. There is no financial or personal relationship with organizations that could potentially be perceived as influencing this research.

References

[1] Van Heerden A, Tomlinson M, Swartz L. Point of care in your pocket: a research agenda for the field of m-health. Bull World Health Organ. 2012;90(5):393-4.

[2] Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, Tsai AC, Jackson D. The use of mobile phones as a data collection tool: A report from a household survey in South Africa. BMC Med Inform Decis Mak. 2009;9:51.

[3] Siedner MJ, Haberer JE, Bwana MB, Ware NC, Bangsberg DR. High acceptability for cell phone text messages to improve communication of laboratory results with HIV-infected patients in rural Uganda: a cross-sectional survey study. BMC Med Inform Decis Mak. 2012;12:56.

[4] Lund S, Rasch V, Hemed M, Boas IM, Said A, Said K, Makundu MH, Nielsen BB. Mobile phone intervention reduces perinatal mortality in zanzibar: secondary outcomes of a cluster randomized controlled trial. JMIR Mhealth Uhealth. 2014;2(1):e15.

[5] Rotheram-Borus MJ, le Roux IM, Tomlinson M, Mbewu N, Comulada WS, le Roux K, Stewart J, O'Connor MJ, Hartley M, Desmond K, Greco E, Worthman CM, Idemundia F, Swendeman D. Philani Plus (+): a Mentor Mother community health worker home visiting program to improve maternal and infants' outcomes. Prev Sci. 2011;12(4):372-88.

[6] Cho MJ, Sim JL, Hwang SY. Development of smartphone educational application for patients with coronary artery disease. Healthc Inform Res. 2014;20(2):117-24.

[7] William MR, Elzie D, Sebuwufu D, Kiguli J, Bazeyo W. Opportunities for strategic use of e-learning in scaling up disaster management capacity in Eastern Africa: a descriptive analysis. East Afr J Public Health. 2013;10(2):403-9.

[8] Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U-V. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth. 2014;2(2):e24.

[9] De la Vega R, Miró J. mHealth: A Strategic Field without a Solid Scientific Soul. A Systematic Review of Pain-Related Apps. PLoS One [Internet]. 2014;9(7).

[10] Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med. 2013;18(3):90-2.

[11] McCartney M. How do we know whether medical apps work? BMJ. 2013;346:f1811.

[12] Cummings E, Borycki EM, Roehrer E. Issues and considerations for healthcare consumers using mobile applications. Stud Health Technol Inform. 2013;183:227-31.

[13] O'Neill S, Brady RR. Colorectal smartphone apps: opportunities and risks. Colorectal Dis. 2012;14(9):530-4.

[14] Visvanathan A, Hamilton A, Brady RRW. Smartphone apps in microbiology--is better regulation required? Clin Microbiol Infect. 2012;18(7):E218-20.

[15] Wolf JA, Moreau JF, Akilov O, Patton T, English JC, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. JAMA Dermatol. 2013;149(4):422-6.

[16] Ferrero NA, Morrell DS, Burkhart CN. Skin scan: a demonstration of the need for FDA regulation of medical apps on iPhone. J Am Acad Dermatol. 2013;68(3):515-6.

[17] Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. J Telemed Telecare. 2012;18(4):235-8.

[18] McKinstry B. Currently available smartphone apps for asthma have worrying deficiencies. Evid Based Med. 2013;18(5):e45-e45.

[19] Powell AC, Landman AB, Bates DW. IN search of a few good apps. JAMA. 2014;311(18):1851-2.

[20] Chan SR, Misra S. Certification of mobile apps for health care. JAMA. 2014;312(11):1155-6.

[21] Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform. 2014;5(3):229.

[22] Powell AC, Landman AB, Bates DW. Certification of mobile apps for health care--reply. JAMA. 2014;312(11):1156-7.

[23] Jacob E, Stinson J, Duran J, Gupta A, Gerla M, Ann Lewis M, Zeltzer L. Usability testing of a Smartphone for accessing a web-based e-diary for self-monitoring of pain and symptoms in sickle cell disease. J Pediatr Hematol Oncol. 2012;34(5):326-35.

[24] Spyridonis F, Gronli T-M, Hansen J, Ghinea G. Evaluating the usability of a virtual reality-based Android application in managing the pain experience of wheelchair users. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:2460-3.

[25] Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, Gerstle JT, Alman B, Hopyan S, Strahlendorf C, Portwine C, Johnston DL, Orr M. Development and testing of a multidimensional iPhone pain as-

sessment application for adolescents with cancer. J Med Internet Res. 2013;15(3):e51.

[26] Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res. 2014;16(4):e104.

[27] Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. Journal of Systems and Software. 2011;84(11):2022-33.

Address for correspondence

Mobin YASINI, DMD Santé, Research and development department, 38 rue Cérès, 51100, Reims, France.

Tel : +33 9 73 19 81 19; Email : mobin.yasini@dmd-sante.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-132

Towards Improving Hypertensive Patients Care: Pervasive Monitoring and Diagnosis Support

Jorge Céspedes[a] and Cynthia Villalba[a]

[a] *Facultad Politécnica, Universidad Nacional de Asunción, Paraguay*

Abstract

Hypertension is the most common chronic condition dealt with by primary care physicians and other health practitioners. It usually has no symptoms, causing a delay in diagnosis. Moreover, around 20% of the global population suffers from "white-coat syndrome", which can lead to misdiagnosing hypertension. When diagnosed, patients find it difficult to constantly monitor their blood pressure to ensure it is within acceptable levels. In this work, we propose a pervasive solution model for ambulatory monitoring of hypertensive patients and for supporting a clinician with the task of diagnosing hypertension. It contributes to the selection of attributes and techniques for assisting hypertension diagnosis, and also to an implementation which dynamically adjusts itself to each patient's average blood pressure.

Keywords:

Pervasive Healthcare; Assisted Diagnosis; Hypertension; Patient Monitoring.

Introduction

Arterial hypertension (HTN), has been globally recognized as the leading chronic disease that causes premature cardiovascular mortality and morbidity [1]. If HTN is properly controlled, cardiac mortality will decrease 49%, while cardiovascular mortality will decrease 62% [1]. Conversely, uncontrolled HTN, according to information from the Panamerican Health Organization (PAHO), could cause terrible health issues, such as myocardial infarction, kidney failure, blindness and/or heart failure [2]. Considering the above mentioned facts, an early diagnosis of the disease and providing mechanisms to maintain blood pressure within acceptable levels, is very important. However, there are issues that need to be addressed. First of all, at least 20% of the world's population suffers from "white coat syndrome" [4]. In the context of HTN, this means that a person's blood pressure readings might be higher in the doctor's office, but once he/she leaves the room, it goes back to normal. Another issue, which is opposite of the previous one, is "masked hypertension". It causes normal blood pressure readings in the doctor's office, but readings at home are in the hypertensive range [5].

Hypertensive patients, find it difficult to constantly monitor their blood pressure. In spite of the existence of well tolerated effective drugs, these are not enough to maintain blood pressure within acceptable levels [6]. It is claimed that around 20 million patients, will suffer a hypertensive crisis at some point of their lives [7].

This work presents a pervasive solution model to support hypertension diagnosis and provide monitoring of hypertensive patients. Supporting diagnosis by automated processing of data retrieved from patients could lead to more accurate diagnoses, while monitoring hypertensive patients could introduce several benefits. First, the patient is notified as his/her blood pressure starts to become abnormal. Second, every reading is stored in a central repository, allowing the clinician to view behavior of the blood pressure of any given patient and whether medication needs to be adjusted or not. Finally, the set of data in the repository, could be used in other research works.

This work is organized as follows. The next section presents recent work related to assisted diagnosis and patient monitoring systems. Then, a description of the proposed solution model is presented. Next, we present the methods used to perform tests and results obtained with them. Finally, a discussion of the results and a conclusion of this work is presented.

Recent Work

Assisted Diagnosis Systems

Most systems performing assisted diagnosis of hypertension use machine learning techniques, extracting features from available patient data or from available laboratory test results [8-11]. Whereas, others use techniques different from machine learning [12, 13], which we refer to as non-machine learning. Most of the works, in the non-machine learning group, ask for input from the user, by means of questionnaires or forms the patients need to fill in.

Currently, clinicians could make use of an ambulatory blood pressure monitoring (ABPM) device to diagnose hypertension when white-coat syndrome is suspected. It is worn by patients for 24 hours and once completed, it is connected to a computer that prepares a report for the patient [14]. A disadvantage of this approach is that it doesn't provide real-time notifications as blood pressure starts to become abnormally high, and that 24 hours might not be a long enough period of time to gather relevant data.

As a final point, we can mention that there are many challenges most work doing assisted diagnosis need to address, such as features to choose, sources of data to extract features from, and deciding whether to use machine learning or non-machine learning techniques, since better results might be obtained by choosing one or the other. This work tackles the challenges related to technique and attribute selection.

Patient Monitoring Systems

In this section, we focus on work performing patient monitoring by using wireless sensors and a smartphone to process data generated by them, that is, work oriented to pervasive monitoring. It is worth mentioning, that not only works monitoring hypertensive patients are considered.

There are studies that used specific sensors and machine learning to identify patient's health condition [15, 16]. Others

allow use of any type of sensors, and apply machine learning techniques to check patient's health status [17-19, 34]. On the other hand, some apply non-machine learning techniques and do not constrain sensors to be used for monitoring [18, 20, 32, 33]. Works dedicated exclusively to monitoring hypertensive patients, apply non-machine learning techniques and some of them combine wearable sensors with sensors installed around the patient's environment [21-24].

Similarly to assisted diagnosis systems, these systems also face some challenges, such as developing a system that can adjust itself to every monitored patient, and one that can provide pervasive monitoring, that is, monitoring without disrupting patients' everyday activities. Other challenges are related to technological issues, medical issues, administrative issues, usability, and validation issues [25, 26]. This work tackles the challenge of providing monitoring by means of an implementation which dynamically adjusts itself to every patient that needs to be monitored.

Solution Model

In Figure 1, a graphical description of the solution model is presented. The model has the following behavior. Sensors transmit data over the air to the patient's smartphone. Once the device gets incoming data, it starts real-time processing to determine whether the patient's current health status is normal or requires attention. If it requires attention, an immediate notification is sent to the patient and clinician in charge. In order to support diagnosis, every piece of data is stored in a main server, where offline processing will take place once a diagnosis is required.

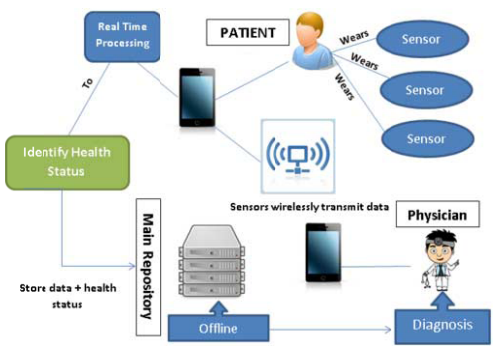


Figure 1 - Graphical description of the solution model

By monitoring a hypertensive patient, we expect to identify situations where severe blood pressure elevations occur. This is done by reading data from an ambulatory blood pressure device, which is attached to patient's left arm. This device measures blood pressure automatically, that is, without explicit request from the user. A smartphone will be in charge of processing the incoming data. Once an abnormal situation is detected, the patient, the clinician treating him or both will receive a notification. The model allows its components to be extended to easily include new sensors; for example, we could add an ECG sensor and include new notifications when emergencies occur.

For assisting hypertension diagnosis, we use machine learning techniques. Although we mentioned many works doing assisted diagnosis in the Recent Work section, there is not a golden rule that establishes which features will lead to reliable diagnosis support. Therefore, this solution proposes use of an activity tracking device, which a patient wears on his wrist, to include sedentary behavior as a new feature, since it has been proven that it is associated with a higher risk of hypertension [27]. Other features are weight, height, BMI, average systolic blood pressure (SBP), and average diastolic blood pressure (DBP).

It is worth mentioning that this is not an attempt to replace the clinician in the task of diagnosing hypertension, nor is it about ruling out standard diagnostic procedures. However, an obvious advantage of this solution model is that it considers longer periods of time, thus leading to more accurate and detailed information [30] which could improve diagnosis. Additionally, it also allows real-time notifications to patients and physicians as readings start to become abnormal. This would allow them to take actions earlier in emergency situations, whereas ABPM would only report abnormal values to the physician once he/she downloads the readings from the device.

Average blood pressure is also used for monitoring, since in order to identify a patient's health condition, the system uses values adjusted to this patient. This is a valid interesting approach since essential hypertension produces structural and functional modifications as a result of the growth or remodelling of blood vessels [28]. We will refer to these values as dynamic values. Previous work has used both dynamic and fixed values to monitor a hypertensive patient [22].

It is important to mention that for calculating average SBP and DBP, we used a weighted average equation (1) that proved to be better than arithmetic mean [24] when certain readings might be deleted due to errors. It considers the number of minutes elapsed between one reading and the next.

$$wAvg = \frac{\Sigma_i (L_i * P_i)}{\Sigma_i P_i} \qquad (1)$$

$$P_i = \mid T_{i+1} - T_i \mid$$

Methods

Test Method for Assited Diagnosis

A prototype of the proposed model was built to perform tests. Even though blood pressure readings can be obtained directly from worn sensors, it is not mandatory to do so. It can also be obtained manually, by requesting a patient to measure his blood pressure and then providing an interface where he can upload the current reading. But this manual approach is less convenient, as the patient can inadvertently make mistakes and has to interrupt his activities to measure his blood pressure. But, this latter approach is convenient when wireless devices are not available, and has been adopted by this work for getting required data. Sixty patients were selected from two Paraguayan hospitals to be part of the tests. Only those 18 – 90 years old, who were able to measure their blood pressure, were enrolled. Patients were requested to write their blood pressure readings down every eight hours, until they completed 10 readings each. Sedentary behavior information, like blood pressure readings, can also be obtained manually by asking patient for information about physical activities performed every day. Once patients accepted to be part of the test, they were interviewed to find out about sedentarism and obtain demographic data. A doctor classified each one of the 60 patients as hypertensive or non-hypertensive (i.e. healthy).

In order to apply a machine learning technique for assisted diagnosis, the full data set was grouped in two sets. The first one contained 35 records and it was used as the "training set", while the latter had 25 records and was used as "test set". Records in the "training set" were partitioned as follows: 20%

of the patients were sendentary and hypertensive, while 8% were hypertensive but non-sedentary. Additionally, 43% of them were non-sedentary and healthy, while 29% were sedentary and healthy. The test set was divided as follows, 40% of patients suffered from hypertension and 60% were healthy. We used an online learning algorithm consisting of a variation of gradient descent, since most work dedicated to assisting diagnoses choose online learning algorithms to generate suggestions [31].

In order to validate the assisted diagnosis systems, a confusion matrix was built. A confusion matrix is a table of at least size m by m, where an entry, CM(i, j) indicates the number of cases i labelled by the system as class j [29]. For the system to have good accuracy, values along the diagonal would be far from zero, while the rest would be close to zero.

The following measures were taken into account [29]:

1. Sensitivity: refers to how well the system can identify a hypertensive patient (True Positive / Positive).
2. Specificity: refers to how well the system can identify a healthy patient (True Negative / Negative).
3. Precision: used to access percentage of hypertensive patients that are actually hypertensive (True Positive / (True Positive + False Positive)).
4. Accuracy: It is a function of sensitivity and specificity (Sensitivity * Positive / (positive + negative) + Specificity * Negative / (positive + negative)).

Where:

- Positive is number of hypertensive patients.
- Negative is number of healthy patients.
- True Positive is number of patients suggested as hypertensive that actually are hypertensive.
- True Negative is number of patients suggested as healthy that actually are healthy.
- False Positive is number of healthy patients suggested as hypertensive patients.
- False Negative is number of patients suggested as healthy but actually suffer from hypertension

Test Method for Monitoring System

The monitoring system is comprised of two main components, a data gathering component and a health condition inference component. For testing purposes, simulated blood pressure readings were used, that is, the data gathering component did not connect to sensors, instead it obtained data from a simulator. The simulator starts with a patient's average SBP and DBP values, and then several events are generated, where each event would alter a patient's blood pressure. An event could be one of the following: **salty food, rough emotions (e.g., anger), physical activity, relaxation, and no change**. Each event, but the last one, causes blood pressure to increase or decrease by three levels: low, medium, high. The health condition inference component calculated average SBP and DBP as new events appeared, using the weighted average formula described in the Solution Model section. However, since values were simulated and none of them was invalid, weighted average and arithmetic median would have lead to the same values.

To perform tests, three fictitious hypertensive patients were created and events were generated all day long every 30 minutes. That is, 48 readings per day with a total of 240 readings for each patient, since these fictitious patients were monitored for five days. Even though blood pressure readings were simulated, we expect to test the system with real patients and sensors in upcoming phases.

The purpose of the test was to observe the differences between using values adjusted to each patient (i.e. dynamic values) and fixed values.

Results

Test Results for Assisted Diagnosis System

In Table 1 we can observe the results obtained after performing the test. The system indicates that 10 patients suffer from hypertension and 12 patients are healthy. It has a false positive rate of three, which means that three patients are considered hypertensive even though they are healthy. It does not generate any false negatives. It has 100% sensitivity and 80% specificity, which has been affected by the existence of false positives. This also affected the obtained precision, which was 77%. Finally, the system has 88% accuracy, which is affected by the specificity the system has.

Table 1 - Values obtained for each measure

Measure	Value
True Positive	10
True Negative	12
False Positive	3
False Negative	0
Sensitivity (%)	100
Specificity (%)	80
Precision (%)	77
Accuracy (%)	88

Test Results for Monitoring System

In Figure 2, we can observe SBP and DBP values for patient P1 during a single day, displaying generated alerts. Dashed orange and solid red lines represent warning and critical alerts for dynamic values, respectively. Whereas, dotted light blue and mixed blue lines represent warning and critical alerts for fixed values, respectively.

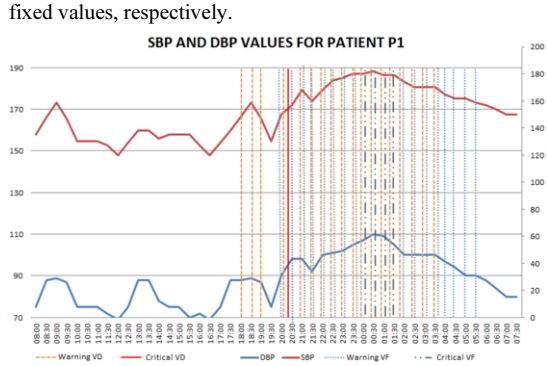


Figure 2 - SBP and DBP values for patient P1 for a day. Warning and Critical alerts are shown for both dynamic values and fixed values.

It is important to note that only one critical alert was generated when dynamic values were used. Though SBP/DBP values higher than the one at the specified time were generated, only warning alerts occurred later. This is due to the average blood pressure of P1 adjusting itself to a new value. However, this might be seen as an incorrect behavior because in practice, average blood pressure cannot change in such a short period of time. Thus, Figure 2 exposes two important matters to be defined when using dynamic values instead of fixed ones, that is, the time window to consider blood pressure readings for

calculating average blood pressure and the frequency at which blood pressure is to be measured. These must be defined so that average blood pressure is realistic.

The results from the tests are displayed in Table 2. We can observe a lower number of warning and critical alarms when dynamic values are used.

Table 2– Results for Monitoring System

Patient	Fixed Values		Dynamic Values	
	Warning Alarms	*Critical Alarms*	*Warning Alarms*	*Critical Alarms*
P1	80	20	76	3
P2	103	92	57	4
P3	120	71	86	5

Discussion

In Table 3, we show the results obtained by work dedicated to assisting hypertension diagnosis along with ours. Information in the table encourages us to believe that we are on the right track since sensitivity and specifity values are greater than those obtained by some recent works. However, a comparison cannot be made since datasets are different. Our results might be improved, perhaps by using larger datasets or by trying different attribute combinations.

Table 3 - Values obtained by others and this proposal

Measure	Su&Wu	Ture et al.	Hsu et al.	Proposal
Sensitivity	46.89%	95.24%	--	100%
Specificity	73.96%	71.79%	--	80%
Accuracy	72%	--	92.8%	88%

The monitoring system obtained results similar to those in [21], where the authors showed that fewer alerts are generated when values adjusted to patients are used. We expect that, in practice, this translates to generating critical alerts only when a patient's current health status needs attention from the clinician in charge.

Conclusion

This work presented a pervasive solution model to support hypertension diagnosis and provide monitoring of hypertensive patients. Although, there are many works on assisting diagnosis, there are challenges that need to be addressed. One of these is the selection of features to be used to accomplish the task. This work introduced a feature that previously mentioned works had not included, that is, whether a patient has sedentary behavior or not. We also contributed to the selection of a diagnosis technique, by choosing an online learning algorithm different from artificial neural networks and support vector machines, which are the most commonly adopted techniques by recent works.

Furthemore, identifying a patient's health condition, by using values adjusted to himself, allowed the system to generate fewer alarms than the case when fixed values were used. In the context of hypertension, this is a valid approach, since a severe blood pressure elevation might represent a crisis for some patients, but it might not be dangerous for someone who often has high blood pressure.

As future work, we expect to try the prototype with real patients. That is, to obtain data from the ambulatory blood pressure sensor and process them in real time instead of simulating readings, since this will allow us to create a confusion matrix; and perform an analysis similar to that in Test Results for Diagnosis System. We also expect to get feedback from doctors and patients to validate the system.

References

[1] Lewanczuk R. Hypertension as a chronic disease: what can be done at a regional level? Can J Cardiol, 2008;24(6):483-484.

[2] Panamerican Health Organization. Hypertension. http://www.paho.org/hipertension/. Accessed: October, 2014.

[4] Sine R. Beyond white coat syndrome: fear of doctors and tests can hinder preventive healthcare. http://www.webmd.com/anxiety-panic/features/beyond-white-coat-syndrome. Accessed: October, 2014.

[5] Papadopoulos DP, Makris TK. Masked hypertension definition, impact, outcomes: a critical review. J Clin Hypertens, 2007;9(12):956-963.

[6] Bernocchi P, Scalvini S, Bertacchini F, Rivadossi F, Muiesan ML. Home based telemedicine intervention for patients with uncontrolled hypertension: a real life non-randomized study. BMC Med Inform Decis Mak, 2014;14(1):52.

[7] Marik PE, Rivera R. Hypertensive emergencies: an update. Curr Opin Crit Care, 2011;17(6):569-580.

[8] Ture M, Kurt I, Kurum AT, Ozdamar K. Comparing classification techniques for predicting essential hypertension. Expert Syst Appl, 2005;29(3):583-588.

[9] Sumathi B, Santhakumaran A. Pre-diagnosis of hypertension using artificial neural network. GJCST, 2011;11(2):43-47.

[10] Das S, Ghosh PK. Hypertension diagnosis: a comparative study using fuzzy expert system and neuro fuzzy system. In: IEEE International Conference on Fuzzy Systems (FUZZ); 7-10 July 2013; Hyderabad, Pakistan. 2013. pp. 1-7.

[11] Su C, Wu C. JADE implemented mobile multi-agent based, distributed information platform for pervasive health care monitoring. Appl Soft Comput, 2011;11(1):315-325.

[12] Hsu K, Chiu C, Chiu N, Lee P, Chiu W, Liu T, Hwang C. A case-based classifier for hypertension detection. Knowl-Based Syst, 2011;24(1):33-39.

[13] Abdullah AA, Zakaria Z, Mohamad NF. Design and development of fuzzy expert system for diagnosis of hypertension. In: Second International Conference on Intelligent Systems, Modelling and Simulation (ISMS); 25-27 January 2011; Kuala Lumpur, Malaysia. IEEE; 2011. pp. 113-117.

[14] National Heart Foundation of Australia. Ambulatory blood pressure monitoring. Test and Results Series. 2011:877-880.

[15] Oresko J. Portable heart attack warning system by monitoring the ST segment via smartphone electrocardiogram. 2010. Master's thesis, University of Pittsburgh.

[16] He Y, Li Y, Yin C. Falling-incident detection and alarm by smartphone with multimedia messaging service (MMS). E-Health Telecommun Syst and Netw, 2012;1(1):1-5.

[17] Lee RG, Chen KC, Hsiao CC, Tseng CL. A mobile care system with alert mechanism. IEEE Trans Inf Technol Biomed, 2007;11(5):507-517.

[18] Otto C, Milenkovic A, Sanders C, Jovanov E. System architecture of a wireless body area sensor network for ubiquitous health monitoring. JMM, 2005;1(4):307-326.

[19] Gund M, Andhalkar S, Patil D, Wadhai VM. An intelligent architecture for multi-agent based m-Health care system. IJCTT, 2011;1(1):1-5.

[20] Su, C, Wu C. JADE implemented mobile multi-agent based, distributed information platform for pervasive health care monitoring. Appl Soft Comput, 2011;11(1):315-325.

[21] Baig MM, GholamHosseini H. A remote monitoring system with early diagnosis of hypertension and hypotension. In: IEEE Point-of-Care Healthcare Technologies (PHT); 16-18 January; Bangalore, India. IEEE. 2013. Pp. 34-37.

[22] Copetti A, Loques O, Leite JCB, Barbosa TPC, da Nóbrega ACL. Intelligent context-aware monitoring of hypertensive patients. In: 3rd International Conference on Pervasive Computing Technologies for Healthcare; 1-3 April 2009; London, United Kingdom. IEEE. 2009. pp. 1-6.

[23] Eungyeong K, Lee M. Multi-agent-based U-healthcare system for managing hypertension. In: International Conference on Convergence Information Technology; 21-23 November 2007; Gyeongju, South Korea. IEEE. 2007. pp. 1694-1699.

[24] Mena L, Felix, VG, Ostos R, Gonzalez, JA, Cervantes A, Ochoa A, Ruiz C, Ramos R, Maestre GE. Mobile personal health system for ambulatory blood pressure monitoring. Comput Math Methods Med, 2013;2013:13.

[25] Lymberis A. Research and development of smart wearable health applications: the challenge ahead. Stud Health Technol Inform, 2004;108:155-161.

[26] Varshney U. Pervasive healthcare and wireless health monitoring. Mobile Netw Appl, 2007;12(2-3):113-127.

[27] Beunza JJ, Martínez-González MA, Ebrahim S, Bes-Rastrollo M, Núñez J, Martínez JA, Alonso A. Sedentary behaviors and the risk of incident hypertension: the SUN cohort. Am J Hypertens, 2007;20(11):1156-1162.

[28] Guyton A, Hall JE. Fisiología y Fisiopatología. McGraw-Hill Interamericana, 1998.

[29] Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3rd ed. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann; 2011.

[30] Wenjin L, Hong Y, Bing T, Zhiping S, Ping W, Wenhui W, Xueqing B, Xia S, Junwei Y. Comparison of 44-hour and fixed 24-hour ambulatory blood pressure monitoring in dialysis patients. J Clin Hypertens, 2014;16(1):63-69.

[31] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform, 2002;35(5-6):352-359.

[32] Fensli R, Gunnarson E, Gundersen TA. Wearable ECG-recording system for continuous arrhythmia monitoring in a wireless telehome-care situation. In: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems; 23-24 June 2005; IEEE. 2005. pp. 407-412.

[33] Moron MJ, Luque JR, Botella AA, Cuberos EJ, Casilari E, Diaz-Estrella A. J2ME and smart phones as platform for a Bluetooth body area network for patient-telemonitoring. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE. 2007. pp. 2791-2794.

[34] Bellos C, Papadopoulos A, Rosso R., Fotiadis DI. A support vector machine approach for categorization of patients suffering from chronic diseases. In: Lin JC. Fotiadis DI. Arredondo Waldmeyer MT. Nikita KS, editors. Wireless Mobile Communication and Healthcare, volume 83 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Berlin, Germany. Springer; 2012. pp. 264-267.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-137

E-Patient Reputation in Health Forums

Amine Abdaoui[a], Jérôme Azé[a], Sandra Bringay[a] and Pascal Poncelet[a]

[a] *LIRMM UM2 CNRS, UMR 5506, 161 Rue Ada, 34095 Montpellier, France*

Abstract

Online health forums are increasingly used by patients to get information and help related to their health. However, information reliability in these forums is unfortunately not always guaranteed. Obviously, consequences of self-diagnosis may be severe on the patient's health if measures are taken without consulting a doctor. Many works on trust issues related to social media have been proposed, but most of them mainly focus only on the structure part of the social network (number of posts, number of likes, etc.). In the case of online health forums, a lot of trust and distrust is expressed inside the posted messages and cannot be inferred by only considering the structure. In this study, we rather suggest inferring the user's trustworthiness from the replies he receives in the forum. The proposed method is divided into three main steps: First, the recipient(s) of each post must be identified. Next, the trust or distrust expressed in these posts is evaluated. Finally, the user's reputation is computed by aggregating all the posts he received. Conducted experiments using a manually annotated corpus are encouraging.

Keywords:

Trust, Reputation, Social media, Online health forums.

Introduction

Internet is allowing patients to play a more active role in their own health care. Today, more than 46% of patients in 12 different countries use internet for self-diagnosis [1]. The use of Web-derived health information is rapidly increasing and has been termed as the *e-patient revolution* [2]. They have a strong desire to learn, to understand their own symptoms and to have access to medical knowledge. Although accessible and easy to use, the reliability of information on Internet represents a major risk on e-patients health. The consequences of an erroneous self-diagnosis are difficult to estimate if measures are taken without consulting a doctor. Indeed, only 21% of e-patients ask their physician confirmation of information obtained from Internet [1]. According to a recent study conducted by the Health On the Net foundation, 90% of e-patients use search engines to initiate their requests. Most of the returned links propose health forums which are used by more than 50% of e-patients. Thus, these forums are becoming the first source of medical information on the Internet. They are areas of exchange where patients, on condition of anonymity, freely relate their personal experiences and give their views and advices.

It is difficult to prevent e-patients to consult irrelevant or unreliable information on health forums; however it is possible to design tools to highlight the trustworthiness of information as well as the users in it. Many online health forums give a rank to each user, which is usually based only on the number of messages posted since his registration. Actually such ranking does not really give a good estimation of users' trustworthiness. A discussion with moderators of a French forum confirmed this intuition, since they know trusted users who post few messages and untrusted users who posted a lot of messages. In a previous work, we proposed a method to automatically distinguish posts made by health experts from those made by laymen [3]. In this new work, we are interested on the reputation of online health forum users independently from their medical roles. Many definitions of trust and reputation can be found in the literature according to each context [4]–[7]. Here we define the trust that a user *A* gives to a user *B* as: "*the belief of A in the accuracy of the information posted by B*", and the reputation of a user *A* as "*the aggregation of the values of trust given to user A*". Most studies of trust in social networks usually focus on the structure of the website (ratings, number of posts, number of likes, number of quotes, distance between posts, etc.) [8]. However, explicit liking and quoting functionalities are rarely used in the case of online health forums. For example, in the French health forum *CancerDuSein.org* only 2% of posts have explicit quoting. Besides, most users in this forum prefer posting a new reply where they express their agreement or recognition rather than simply pressing the like button.

In this study we suggest to infer the user's reputation from the replies he receives in the forum. Therefore, our method is divided into three main steps. First, links between each post and the person(s) to whom it is addressed, also called the recipient(s), are identified. Some works have already addressed this task [9], [10]. Three types of relationships have been extracted: structural relationships, name relationships and text quotation relationships. In this work, we consider nine kinds of different relationships. After that, posts are evaluated and classified into one of the following classes: trust, distrust and neutral, according to the use of agreement, disagreement and thanking expressions. Finally, the user's reputation is computed by aggregating the trust and distrust expressed in the posts he received. It may computed in the whole forum or in specific topics by aggregating only replies posted in that specific topic. Users' reputations can be used either by moderators in order to investigate users having very bad reputations, or by forum readers in order to have an idea about the authors trustworthiness.

The rest of the paper is organized as follows. Section 2 presents the corpus and describes the used methods. Section 3 presents the obtained results and Section 4 discusses them. Finally, Section 5 concludes and gives our main perspectives.

Materials and Methods

First, the corpus of study is described. Then, the three main steps of our method are presented: identifying the recipient(s) of each post, inferring the trust expressed by each post and computing the reputation of each user.

Corpus of study

CancerDuSein.org is a French health forum specialized in breast cancer. 1,050 threads have been collected which holds 16,961 messages posted by 675 users. It represents all the data that have been posted between October 2011 and November 2013. Some threads have more than 500 posts, which make the use of semi-automatic systems a challenging task. This forum gathers women with breast cancer or their families, who want to exchange their experiences, advices and emotional support. However, the consequences of acting on incorrect advices can be severe. The forum gives a rank to each user based on the number of posts since his registration as described in Table 1. However, an active member on the forum is not necessarily a trusted member and similarly a new member is not necessarily an untrusted one.

Table 1 – Number of posts for each rank

Rank	Number of posts
New member	[0, 20[
Regular member	[20, 40[
Accustomed member	[40, 80[
Active member	>80

Step 1: Finding the recipient(s)

The first step of our method consists in finding the recipient(s) of each post in the forum. In order to construct a network of replies, a rule based heuristic has been developed. Nine rules have been designed and applied chronologically as described below. If a message does not match the first rule, the heuristic will check the second one and so on. The first post in each thread is not considered since it does not answer anybody.

Explicit quoting: *CancerDuSein.org* allows users to explicitly quote another user's post. However, only 2% of posts on the Website contain explicit quoting. These quotes have been detected using the HTML tag <*quote*> and by comparing the content of these tags and the pseudonym of the quoted user with the messages posted before in the same thread. This process allowed us to detect 312 quotes automatically. The rest of quotes (37) have been related manually because the quoted text has been modified or truncated by the user.

Second posts: Messages posted at the second place in each thread have been considered as replying to the first one.

Names and pseudonyms: If a message contains the pseudonym or the name of a user who previously posted a message in the same thread, then this user is considered as the recipient of the message. The pseudonyms have been extracted automatically while the names have been extracted from the signatures and validated manually. The following preprocessings have been applied in order to detect names and pseudonyms inside the text:

- Remove all non-alphabetic characters except spaces (***John Woe 34*** becomes *John Woe*).
- Replace all accented characters by the corresponding non-accented ones (*Jérôme* becomes *Jerome*).
- Lowercasing (*Sandra* becomes *sandra*).

Grouped posts: If a message contains a group marker ("hello everyone", "Hi girls", "Thank you all", etc.) then all the users who previously posted in the same thread are considered as recipients for this post.

Second person pronouns: In French, singular second person pronouns and plural second person pronouns are different. If a singular second person pronoun is used then the recipient is the author of the previous post.

Activator posts: When the activator of the thread (the user who opened the the thread by posting the first message) posts a new message in the same thread, we consider that all the users who posted after his last message are recipients of his new message.

Questions: If the message contains a question then the message is addressed to all the users who previously posted in the same thread.

Answers: If there is a question posted before in the thread, the recipient is the user who posted this question.

Default: If none of the rules mentioned before is satisfied, we consider that the recipient of the message is the activator.

Table 2 presents the number of posts that match each rule.

Table 2 – Number of posts that match each rule

Rule	Number of posts
Explicit quoting	349
Second posts	942
Names and pseudonyms	7,121
Grouped posts	740
Second person pronouns	2,406
Activator posts	1,239
Questions	298
Answers	1,790
Default	772
Total	**15,657**

Step 2: Inferring the trust

The second step consists in classifying each post according to the expressed trust. Posts containing agreement and thanking expressions have been considered as expressing trust to the answered person. Posts containing disagreement expressions have been considered as expressing distrust. The rest of posts have been considered as neutral.

Building the lists of expressions: The expressions of agreement, disagreement and thanking have been built manually based on terms extracted from the LAROUSSE thesaurus [11] and with the help of some specific Websites. All these expressions have been lemmatized in order to detect all the forms of words. Finally, the trust list contains 34 expressions of agreement and thanking while the distrust list contains 15 expressions of disagreement.

Negation: if one of the trust expressions is under the scope of a negation term, it is considered as a distrust expression and vice versa. The scope of a negation term may be two words after, two words before or two words after and two words before according to the nature of the negation term.

Computing the frequencies and classifying the posts: First, all posts have been lowercased, lemmatized using the TreeTagger tool [12] and corrected using the spell checker

Aspell (www.aspell.net). Next, posts are classified as expressing trust, distrut or neutral as follows:

- If a post contains more trust expressions than distrust ones, it is considered as expressing trust.
- If a post contains more distrust expressions than trust ones, it is considered as expressing distrust.
- Otherwise, it is considered as neutral.

Step 3: Computing the reputation

Once the posts addressed to each user are identified and the trust expressed in each post evaluated, we can compute the reputation of each user based on the trust and/or distrust expressed by the replies he received. We suggest computing the difference between the rate of replies expressing trust and the rate of replies expressing distrust. For a user "*u*" the reputation is computed as follows:

$$Reputation(u) = \begin{cases} \frac{NRT(u) - NRD(u)}{NR(u)}, if\ NR(u) \neq 0 \\ 0 \quad\quad\quad\quad\quad\quad , Otherwise \end{cases}$$

Where:

NRT(*u*) is the number of replies expressing trust to user "*u*"
NRD(*u*) is the number of replies expressing distrust to user "*u*"
NR(*u*) is the total number of replies addressed to user "*u*"

Reputation(*u*) belongs to the range [-1, 1]

Results

In order to evaluate the rule based heuristic that finds the recipient(s) of each post and the automatic inference of trust, 2,433 manual annotations have been done. The results obtained using these annotations and those obtained after computing the reputation are described below.

Step 1: Finding the recipient

Two datasets have been used to test our rule based heuristic. The rules have been designed according to the development set (the authors used this dataset to have an idea on the rules that need to be designed). After that, a test set has been used to test our heuristic on unseen threads.

Table 3 – Number of threads, posts and links found by the heuristic in each dataset

Datasets	Number of threads	Number of posts	Number of links
Development set	10	105	152
Test set	10	109	150

Prior-assessment: 15 non-expert annotators (they do not know the designed rules) annotated our two datasets. Each one annotated between 1 and 5 threads so that each thread had 3 annotators. The goal was to find the recipient(s) of each post without knowing the results of our heuristic. Classical measures of agreement are not well adapted, here we simply present the number of links (message, recipient) found by our three annotators at the same time, the number of links found by two out of the three annotators and the number of links found by only one annotator.

Table 4 – Links found by one, two and by the three annotators in each dataset

Datasets	Found by	Number of links	Percentage
Development set	3 annotators	102	53.4%
	2 annotators	54	28.3%
	1 annotator	35	18.3%
	Total	191	100%
Test set	3 annotators	103	60.3%
	2 annotators	44	25.7%
	1 annotator	24	14%
	Total	171	100%

Post-assessment: 3 expert annotators (the authors) annotated the links found by the heuristic in the two datasets. The goal was to validate or not the links found automatically with the possibility of adding a link which was not found by the heuristic. The agreement between the annotators was very good, (the obtained Fleiss' Kappa [13] is **0.89** for the development set and **0.74** for the test set).

Evaluation: Using these annotations, the quality of the developed heuristic has been evaluated. The links obtained automatically have been compared with those obtained from the annotations by considering only those that have been validated by more than two annotators (a majority vote). Table 5 presents the obtained precision, recall and F1-score.

Table 5 – Precision, recall and F1-score of the heuristic obtained on both dataset using prior and post assessments

		P	R	F
Prior assess-ment	Development set	0.70	0.68	0.69
	Test set	0.81	0.83	0.82
Post assess-ment	Development set	0.80	0.84	0.82
	Test set	0.83	1	0.91

Step 2: Inferring the trust

Two new datasets have been used to evaluate the automatic trust inference. Unlike the first step where both datasets had prior and post assessment, here prior-assessment has been done only for the first dataset and post-assessment has been done only for the second one.

Prior-assessment: 3 annotators annotated the trust expressed in 97 messages without knowing the results of the automatic system. The agreement between them was less than the recipient assessment but still acceptable (the obtained Fleiss' Kappa is **0.61**).

Post-assessment: The same 3 annotators annotated the trust expressed in 102 other messages. The results of the automatic system have been displayed, and annotators can chose the same value of trust or another one. The agreement between the annotators was also acceptable (the obtained Fleiss' Kappa is **0.69**).

Evaluation: The results obtained by comparing the classification made by the system with the annotations are presented Table 6. The annotations have been combined by using a majority vote.

Table 6 – Precision, recall and F1-score of the trust inference system using prior and post assessments

Datasets	Class	Number of posts	P	R	F
Prior assessment	Trust	28	0.67	0.93	0.78
	Distrust	4	0.50	0.25	0.33
	Neutral	65	0.96	0.83	0.89
	Global	**97**	**0.86**	**0.84**	**0.84**
Post assessment	Trust	31	0.77	0.87	0.82
	Distrust	5	0.23	0.60	0.33
	Neutral	64	0.92	0.75	0.83
	Global	**100**	**0.84**	**0.78**	**0.80**

Step 3: Computing the reputation

First, Table 7 presents the number of authors, the mean of the reputation and the standard deviation of the reputation for each rank.

Table 7 - The mean and the standard deviation of the reputation for each rank

Rank	Number of authors	Avg(R)	Std(R)
New member	561	0.29	0.29
Regular member	42	0.35	0.10
Accustomed member	26	0.35	0.09
Active member	41	0.38	0.07

Next, Figure 1 presents a scatterplot of the reputation and the number of posts while Figure 2 presents a scatterplot of the reputation and the number of replies.

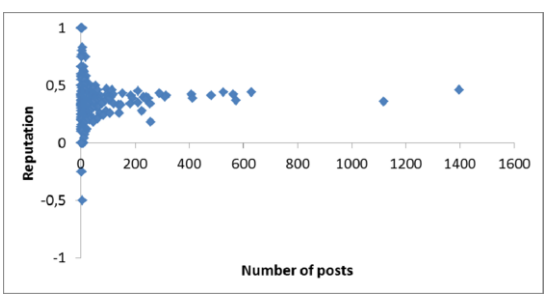


Figure 1 - The reputation and number of posts scatterplot

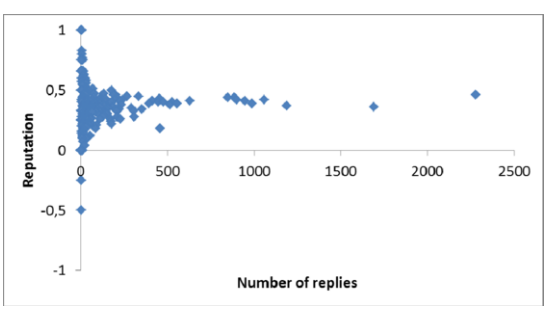


Figure 2- The reputation and the number of replies scatterplot

Discussion

In this section, we discuss the results obtained in each step.

Step 1: Finding the recipient

As expected, results obtained by using a post-assessment have been better from those obtained using prior-assessment. The difference is larger between the recalls than between the precisions. The gain in precision is 0.10 for the development set and 0.02 for the test set whereas the gain in recall is 0.16 for the development set and 0.17 for the test set. This observation can be explained by two reasons. First, the nature of the prior assessment itself gives the non-expert annotators much more freedom to choose the links, which increases the chances of validating links that the heuristic will not find. Furthermore, the non-expert annotators do not know the heuristic rules at all, they may validate links that the heuristic is not designed to find. Their annotations may be useful to add or update some rules. But surprisingly, the results obtained from the test set have been better than those obtained from the development set. Luckily, the test set has less particular cases that we did not implement in our heuristic.

Step 2: Inferring the trust

Unlike the first step, results obtained by using a prior assessment have been slightly better than those obtained using a post assessment. This observation tends to reduce the effect of the knowing the system's results while annotating in the case of trust inference. This small difference may be due to the chosen posts and not to the way that annotations have been done. The results obtained on the trust class have been good, but the recall is higher than the precision using both prior and post assessment. It means that the system finds the majority of posts expressing trust but also gives some posts that are not expressing it as so. Therefore, even if the list of trust expressions has been built manually, it seems to be sufficient to find the majority of trust posts. The results obtained on the neutral class have also been good, but the precision is higher than the recall. It means that the majority of neutral posts have been correctly classified but some posts have not been found by the system (classified in other classes). Finally, the results obtained on the distrust class have been the worst but it is difficult to make conclusions since very few distrust posts have been annotated. In fact, users in this forum do not usually express a lot of disagreement since the first goal is to exchange emotional support.

Step 3: Computing the reputation

Since the forum has very few disagreement posts, the reputation of almost all users is positive. Only, 3 users had negative reputation (two points are superposed in Figure 1 and 2). They posted less than 7 posts and received less than 5 replies. Indeed, the more posts a user receives the more chances he has to receive trust posts and more importantly neutral posts. This is why the standard deviation of the reputation decreases with the increase of the number of posts. Because neutral posts have a mitigating effect to both good and bad reputations.

As presented Table 7, the majority of users have the first rank on the forum (New member). The users' reputations mean increases slightly from lower rank levels to higher ones while the standard deviation decreases. However, authors from the first rank level have reputations ranging from -0.5 to 1 which is very diverse. This observation confirms our hypothesis that the Website rank is not a good estimation of the users' trustworthiness.

Conclusion

In this paper, we presented a method to infer the user's reputation from the posts he receives in online health forums. Indeed, a lot of trust is expressed inside the posted text which can not be detected by structure based trust models. Our method may be used either by the users to have an idea on the trustworthiness of each user or by the moderators, for example to reward the users having the best reputations and to detect the ones having the worst reputation. The method is divided into three steps. First, the recipients of each post are identified using a rule based heuristic. Next, the trust expressed in each post is evaluated by searching the use of agreement, disagreement and thanking expressions. Finally, the reputation of each user is obtained by aggregating the trust or distrust expressed in all the posts addressed to him. The method has been tested on one French health forum specialized in breast cancer where users exchange a lot of emotional support but few disagreement. Therefore most of them have got positive reputations. Manual annotations have been done in order to evaluate the methods. The results obtained for the first and the second step are encouraging.

This paper presents a first implementation and test of a method that infers the trust from the posts addressed to each user. Many perspectives can be done in order to go further in this idea. First, the users' reputations are now computed in the whole forum, we may also compute the reputation in each topic since one user may be trusted differently in several topics (reputation is topic dependent), which can be done by considering only posts received for specific topics. Second, even if the annotated corpus is relevant, still now its size is quite small according to the size of the whole forum. Improving a more complete annotation could probably provide more accurate results. Therefore, we are planing to use crowdsourcing services in order to annotate the whole forum. The quality of annotations obtained using crowdsourcing services are usually questioned but many solutions may be used to overcome this issue. For example, we can put the posts that had a perfect agreement in this study between the posts of the new corpus to make sure that the future annotators will correctly do their work. A large annotated corpus will not only gives us a better estimation of the method's performances, but can also be used in order to learn models that will be able to better classify each post according to the expressed trust. Text mining and supervised classification techniques might give better results in this case.

Moreover, following the evolution of the user's reputation since his registration may also be very interesting. Indeed, we noticed that new users usually receive less trust replies than old ones. But, their reputation can increase over time. Especially in the case of chronic diseases, users remain on the forum for years. Using their experiences and the knowledge acquired, they will become experts and will receive more thanking and agreement replies. Finally, the posts expressing trust posted by users having good reputation may have more weight than posts expressing trust posted by users having a bad reputation. One way to include these propagation aspects is to use PageRank based trust models [14]. This algorithm ranks webpages according to their importance. The basic idea is to give more importance to web pages that are pointed by many other important pages. It has been wihdly applied in social media for example to find key users in terms of connectivity and communication activity [15].

Acknowledgement

This paper is based on studies supported by the "Maison des Sciences de l'Homme de Montpellier" (MSH-M) within the framework of the French project "Patient's mind" (https://www.lirmm.fr/patient-mind/pmwiki/pmwiki.php?n=Site.Accueil).

References

[1] H. S. Wald, C. E. Dube, and D. C. Anthony, "Untangling the Web—The impact of Internet use on health care and the physician–patient relationship," *Patient Educ. Couns.*, vol. 68, no. 3, pp. 218–224, Nov. 2007.

[2] S. M. Akerkar and L. S. Bichile, "Doctor patient relationship: changing dynamics in the information age," *J. Postgrad. Med.*, vol. 50, no. 2, pp. 120–122, Jun. 2004.

[3] A. Abdaoui, J. Azé, S. Bringay, and P. Poncelet, "Predicting Medical Roles in Online Health Fora," presented at the 2nd International Conference on Statistical Language and Speech Processing, SLSP 2014, Grenoble, 2014.

[4] M. Deutsch, "Cooperation and trust: Some theoretical notes," in *Nebraska Symposium on Motivation, 1962*, Oxford, England: Univer. Nebraska Press, 1962, pp. 275–320.

[5] S. P. Marsh, "Formalising Trust as a Computational Concept," Ph.D. Thesis, University of Stirling, Department of Mathematics and Computer Science, 1994.

[6] R. Falcone and C. Castelfranchi, "Trust and Deception in Virtual Societies," C. Castelfranchi and Y.-H. Tan, Eds. Norwell, MA, USA: Kluwer Academic Publishers, 2001, pp. 55–90.

[7] J. Golbeck, "Trust and Nuanced Profile Similarity in Online Social Networks," *ACM Trans Web*, vol. 3, no. 4, pp. 12:1–12:33, Sep. 2009.

[8] F. Skopik, H.-L. Truong, and S. Dustdar, "Trust and Reputation Mining in Professional Virtual Communities," in *Proceedings of the 9th International Conference on Web Engineering*, Berlin, Heidelberg, 2009, pp. 76–90.

[9] A. Gruzd and C. Haythornthwaite, "Automated Discovery and Analysis of Social Networks from Threaded Discussions. Paper presented at," in *the International Network of Social Network Analysts, St. Pete Beach*, 2008.

[10] M. Forestier, J. Velcin, and D. Zighed, "Extracting Social Networks to Understand Interaction," in *2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2011, pp. 213–219.

[11] D. Péchoin, *Thésaurus*. Paris: Larousse, 1999.

[12] H. Schmid, *Probabilistic Part-of-Speech Tagging Using Decision Trees*. 1994.

[13] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971.

[14] J. Caverlee, L. Liu, and S. Webb, "Towards Robust Trust Establishment in Web-based Social Networks with Social trust," in *Proceedings of the 17th International Conference on World Wide Web*, New York, NY, USA, 2008, pp. 1163–1164.

[15] J. Heidemann, M. Klier, and F. Probst. Identifying Key Users in Online Social Networks: A PageRank Based Approach. *ICIS 2010 Proc.*, Jan. 2010.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-142

Web-based Self-management Support Interventions for Cancer Survivors: A Systematic Review and Meta-analyses

Ae Ran Kim[a], Hyeoun-Ae Park[a,b]

[a]College of Nursing & Systems Biomedical Informatics Research Center, Seoul National University, Seoul, Korea
[b]Research Institute of Nursing Science, Seoul National University, Seoul, Korea

Abstract

Those who are living with cancer as a chronic disease need to self-manage the late and long-term effects of cancer and its treatment. We conducted systematic searches of English-language peer-reviewed publications in PubMed, Cumulative Index for Nursing and Allied Health Literature, Cochrane Central Register of Controlled Trials, and EMBASE between January 2000 and June 2014. We searched for web-based interventions designed to help cancer survivors manage their symptoms and the side effects of cancer treatments, which yielded 37 studies that were systematically reviewed. For the meta-analyses, five articles were selected for fatigue, seven for depression, five for anxiety, and five for overall quality of life. The most popular mode of intervention delivery was "peer-to-peer access" in the communicative functions category, followed by "the use of an enriched information environment" in the automated functions category. The effectss across all outcome measures were small to moderate compared to standard care. Healthcare providers could use information technologies to support self-management among cancer survivors based on their needs across the cancer care continuum.

Keywords:
Self-management; Cancer survivor; Web; Intervention.

Introduction

Patients living with a chronic disease such as cancer experience significant limitations with respect to activities of daily living, reduced quality of life, and increased medical costs and burden [1, 2]. People with chronic diseases need to be active participants in their own care and manage their health over time [3]. It is also very important for nurses to be involved in the care of patients with a chronic disease to support and educate them on self-management of the disease.

The number of cancer survivors continues to increase due to improvements in early detection and treatments [4]. Although most chronic cancers are still difficult to cure, they can be managed and controlled as chronic diseases with long-term surveillance and, in some cases, treatment [5]. Since more active participation is required in day-to-day self-management for patients with a chronic disease [6], it is important for cancer survivors to be able to manage the late consequences of treatment or comorbidities on their own [7]. Self-management refers to "the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent to the life with a chronic condition" [8], as guided by healthcare practitioners [9].

Rapid advances in information technology have led to its increasing use among patients with chronic diseases and among healthy people [10]. Information technology can provide self-management tools and material, and can support dynamic decision-making as per patients' needs, preferences, and values [11]. Cancer patients use the Internet to access self-care information regarding symptom management, and are empowered through the decision-making and peer support [12]. The development of more effective self-management interventions for cancer survivors using information technology requires exploration of previous intervention studies for cancer patients and identification of the factors required for effective cancer care [10].

This systematic review is aimed at identifying the general characteristics of web-based self-management support interventions for cancer survivors and to perform the corresponding meta-analyses to assess the effects of these interventions.

The main research questions were as follows:

1. What are the characteristics of the current web-based self-management support interventions for cancer survivors?
2. What modes of intervention delivery are used for cancer survivors on the Web?
3. Were the web-based interventions for cancer survivors more effective than the standard interventions?

Methods

We performed a systematic review and meta-analyses following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-analysis [13]. The following inclusion criteria were applied to select studies for the systematic review:

1. Study subjects: patients diagnosed with cancer or their caregivers.
2. Interventions: Web-based self-management support interventions.
3. Outcome variables: behavioral or health outcomes.
4. Research design: randomized controlled trials or quasi-experimental designs.
5. Publication types: Original research articles published in English in peer-reviewed journals between January 2000 and June 2014.

Studies meeting the following criteria were subjected to further meta-analyses:

1. Studies comparing the effects of Web-based self-management support interventions with those of standard care.

2. Outcome variables reported for more than five studies.
3. Outcomes evaluated with the same measurement tools in more than three studies.

We performed an electronic search in the following four databases: PubMed, Cumulative Index for Nursing and Allied Health Literature, Cochrane Central Register of Controlled Trials, and EMBASE. The following keywords were used for the literature search: "cancer," Web-based ("Internet" OR "Web" OR "online"), self-management interventions ("self-management" OR "self-care" OR "self-guided" OR "self-help" OR "intervention" OR "program"), and study design ("randomized controlled trial" OR "non-randomized controlled trial" OR "quasi-experimental study" OR "before-after study"). The literature search was conducted in July 2014.

For the systematic review, the characteristics of each study were collected: design, subjects, phase, intervention duration, sample size, theory or models used, modes of intervention delivery, and outcome measures. Modes of intervention delivery were classified using the following coding scheme developed by Webb et al. [14]: automated functions, communicative functions, and the use of supplementary modes. Frequencies of the general characteristics and the modes of intervention delivery were tallied among the studies.

The risk of bias in individual studies included in the systematic review was assessed using the Cochrane risk of bias assessment tool [15]. The statistical details about the sample size and outcome variables were collected for the treatment and control groups from the studies included in the meta-analyses using Comprehensive Meta-Analysis Version 2.0. Since some of the individual studies had small samples (for example, n=18 for 2 groups), Hedges' g was used as an effect size [16]. Heterogeneity between studies was evaluated using Q and I^2 statistics. Fixed-effect and random-effect models were used to aggregate the effect sizes of the homogeneous and heterogeneous studies, respectively. Effect sizes of <0.2, 0.2–0.8, and >0.8 are considered to be small, moderate, and large, respectively [17]. The risk of bias across the studies included in meta-analyses was assessed by evaluating the symmetry of funnel plots.

Results

In total 37 articles were selected for the systematic review, and the meta-analysis included 5 articles for fatigue, 7 for depression, 5 for anxiety, and 5 for overall quality of life (Figure 1). The risk of bias in the individual studies included in the systematic review was assessed for the 34 randomized controlled studies. The random sequence generation was adequately described in 30 studies (88.2%), and an allocation concealment procedure was clearly reported for 12 studies (35.3%). Blinding of participants and researchers was described in only three studies (8.8%), but assessment of blinding outcome was not reported at all. However, most of the studies (82.4%) measured outcomes using self-reported questionnaires, and blinding outcome assessment was not applicable to these studies. All pre-specified outcomes were available for all but one of the studies. Thirty-two studies (94.1%) adequately addressed "other sources of bias" (Figure 2).

Thirty different interventions were used in the 37 studies. Regarding the modes of interventions delivery, automated functions were used in 7 interventions, communicative functions in 7 interventions, and both automated and communicative functions in 16 interventions. Supplementary modes were used in 16 interventions (Table 2). The most widely used function in the automated functions category was "the use of an enriched information environment" (n=17), and the most widely used function in the communicative functions category was "peer-to-peer access" (n=18). Examples of the use of an enriched information environment are linking to publically available health recommendations, websites, and multimedia such as videos or audios. Examples of peer-to-peer access are online discussion forums, bulletin boards, and chat rooms. Examples of supplementary modes are use of email, short message service, and telephone.

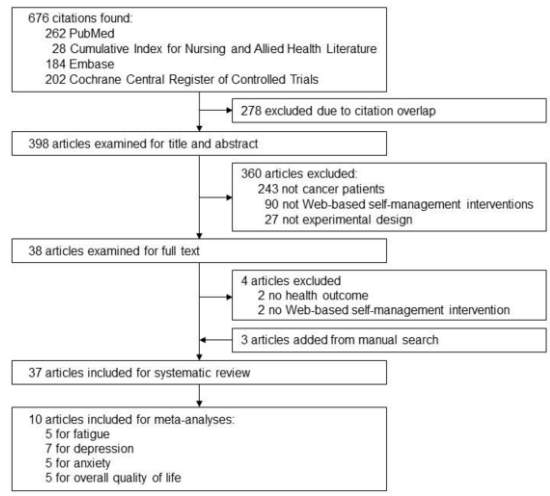


Figure 1 – Analytical framework for the systematic review and meta-analyses.

Table 1 – Summary of the general characteristics of selected studies (n=37)

Charac-teristic	Categories	*n* (%)
Year of publication	2000–2005	3 (8.1)
	2006–2010	6 (16.2)
	2011–2012	8 (21.6)
	2013–2014	20 (54.1)
Study design	Randomized controlled trial	34 (91.9)
	Two groups	29 (78.4)
	More than two groups	5 (13.5)
	Quasi-experimental study	3 (8.1)
	One group	3 (8.1)
Type of disease	Breast cancer	15 (40.6)
	Gynecologic cancer	3 (8.1)
	Hematologic cancer	2 (5.4)
	Prostate cancer	2 (5.4)
	Lung cancer	2 (5.4)
	Pediatric cancer	1 (2.7)
	Breast and prostate cancer	1 (2.7)
	Breast and gynecologic cancer	1 (2.7)
	More than two types of cancer	10 (27.0)

Study phase	During treatment	2 (5.4)
	Post-treatment	15 (40.5)
	From diagnosis to post-treatment	2 (5.4)
	No specific study period	18 (48.6)
Total sample size	179.4±198.7 (mean±SD), range 18~794	
Intervention period (weeks)	28.2±27.9 (mean±SD), range 4~100	
Theoretical basis[1]	Theory-based interventions	16 (43.2)
	Social cognitive theory	5 (13.5)
	Transtheoretical model	3 (8.1)
	Transactional model of stress and coping	2 (5.4)
	Theory of empowering knowledge	2 (5.4)
	Behavioral change model for Internet interventions	1 (2.7)
	Quality health outcomes model	1 (2.7)
	Representational approach	1 (2.7)
	Supportive accountability	1 (2.7)
	Self-determination theory	1 (2.7)
	Theory of informed and shared decision-making	1 (2.7)
	Not-theory-based interventions	21 (56.8)

Synthesis of Results

The effect sizes for fatigue (n=5), depression (n=7), anxiety (n=5), and overall quality of life (n=5) are presented in Figure 3. Since all of the selected study outcomes were homogeneous between studies (fatigue: Q=5.670, p=0.225, I^2=29.458; depression: Q=7.674, p=0.263, I^2=21.809; anxiety: Q=1.416, p=0.841, I^2<0.001; overall quality of life: Q=2.125, p=0.713, I^2<0.001), effect sizes were computed using a fixed-effect model. Web-based self-management support interventions had statistically significant effects on improvements in fatigue (Hedges' g=–0.259, 95% CI: –0.405, –0.114, p<0.001), depression (Hedges' g=–0.169, 95% CI: –0.282, –0.055, p=0.003), anxiety (Hedges' g=–0.293, 95% CI: –0.465, –0.122, p=0.001), and overall quality of life (Hedges' g=–0.221, 95% CI: –0.358, –0.084, p=0.002). Some degree of asymmetry in the funnel plots was observed for the meta-analyses for fatigue and depression, representing the risk of bias across studies.

Table 2 – Web-based modes of delivery[1] (n=30)

Category	Examples	*n* (%)
Automated functions	The use of an enriched information environment	17 (56.7)
	Automated follow-up messages	11 (36.7)
	Automated tailored feedback	8 (26.7)
Communicative functions	Peer-to-peer access	18 (60.0)
	Access to an advisor	10 (33.3)
Supplementary modes	Email	13 (43.3)
	Text message (SMS)[2]	3 (10.0)
	Telephone	2 (6.7)

Discussion

We conducted a systematic review and meta-analyses of studies on the Web-based self-management support interventions for the cancer survivors. Our review revealed that only two theories were used by these studies to develop the intervention: social cognitive theory and the trans-theoretical model (TTM). Previous review studies on Internet-based interventions designed to alter behavior [14] also found that social cognitive theory and the TTM were the two most popular theories used. A behavioral theory can provide a mechanistic basis for specific behavioral changes [14], leading to a theory-based intervention [18]. It would be interesting to determine whether theory-based interventions produce better

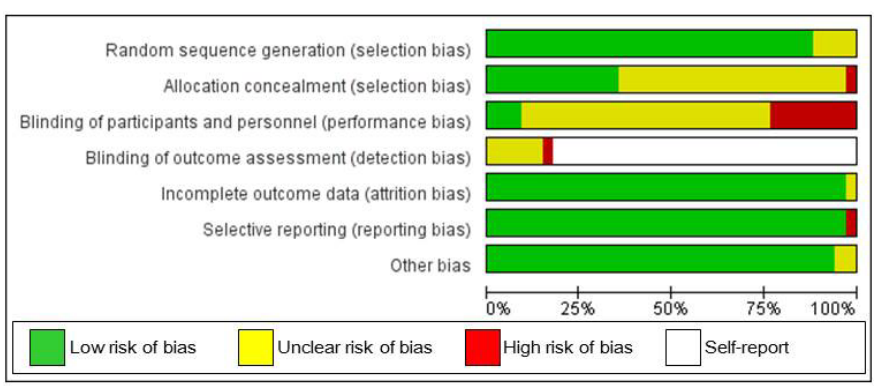


Figure 2 – Risk of bias in individual studies

[1] Multiple counts were possible for one study.

[2] Short message service

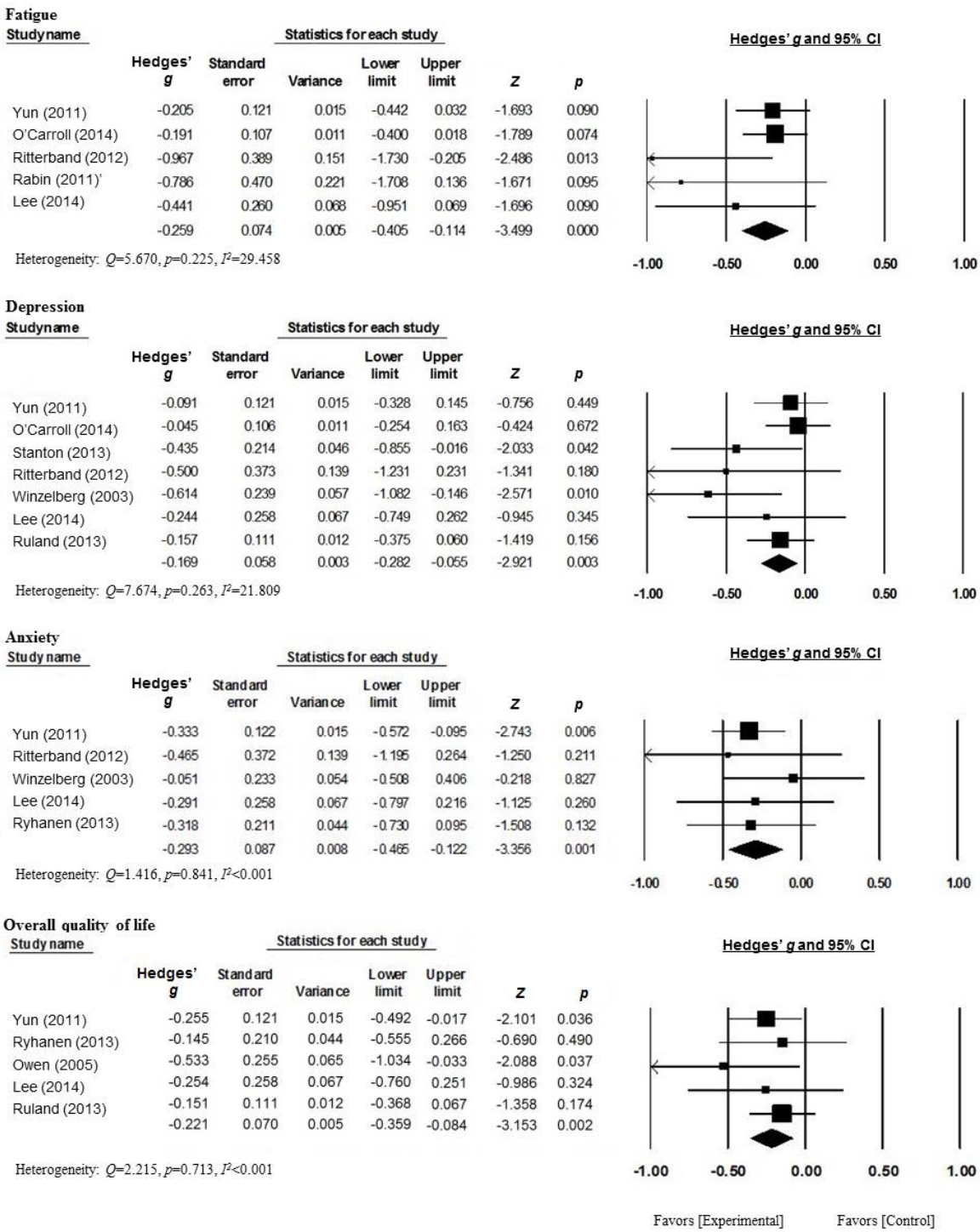


Figure 3 – Effect sizes of fatigue, depression, anxiety, and overall quality of life

behavioral changes compared to interventions that are not theory-based.

The use of an enriched information environment was the most popular function in the automated functions category of intervention delivery modes. Compared to conventional interventions the web-based interventions provide diverse health information in various forms by providing links to rich information sources. Another unique feature of automated functions of the web-based interventions is the tailored feedback, which was used in the eight interventions in the

studies reviewed herein. The automated tailored feedback provides customized information based on specific patient data [19], thereby avoiding the problem of information overload for patients [20]. This will also lead to an active patient participation in self-monitoring.

Peer-to-peer access was the most popular function in the communicative functions category followed by access to an advisor. According to a review study of off-line self-management interventions seeking advice from healthcare providers was the main communicative function [5, 21]. By participating in the online support groups, such as chat rooms and message boards, cancer survivors can compare their disease process with those of their peers, seek opinions from other patients, and learn daily life management skills from other patients [12]. However, the quality of the health information posted by the online support groups cannot be relied on[22]. Thus, it is important for healthcare providers to ensure the reliability and accuracy of information being exchanged and shared in online support groups.

In the meta-analyses, small-to-moderate effects of Web-based self-management support interventions were found on fatigue, depression, anxiety, and overall quality of life as compared to the standard of care. However, these meta-analyses have some limitations. First, the outcome variables were mostly self-reported by patients, which may carry reporting bias [20]. Second, blinding of participants and researchers was reported for only two of the studies (20.0%) included in the meta-analyses, which may have resulted in overestimation of the effect sizes [23]. Finally, we could not test for the existence of publication bias in the meta-analyses since quantifying the risk of error by interpreting the funnel plots was not feasible due to the small number of articles included in the meta-analyses. To the best of our knowledge this is the first study to compute the effects of web-based self-management support interventions on the patient outcomes such as fatigue, depression, anxiety, and overall quality of life for cancer survivors.

Conclusion

Web-based self-management support interventions for cancer survivors were found to be effective in improving fatigue, depression, anxiety, and overall quality of life, with the benefits of the automated and communicative functions available on the Web. Healthcare providers can use these findings to encourage cancer survivors to participate in Web-based self-management support programs. We also recommend that healthcare providers develop intervention strategies that utilize information technologies based on theoretical frameworks to support self-management for cancer survivors along the cancer-care continuum.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028631).

References

[1] Centers for Disease Control and Prevention (CDC). Chronic Diseases, The Power to Prevent, The Call to Control. At a glance 2009 [accessed November 25, 2014]. Available from: http//www.cdc.gov/chronicdisease.

[2] Australian Institute of Health and Welfare (AIHW). Risk Factors Contributing to Chronic Disease [accessed December 2, 2014]. Available from: http://www.aihw.gov.au/Publication-detail/?id=10737421466.

[3] Coster S, Norman I. Cochrane reviews of educational and self-management interventions to guide nursing practice: a review. Int J Nurs Stud. 2009;46(4):508–28.

[4] Jefford M, Karahalios E, Pollard A, Baravelli C, Carey M, Franklin J, et al. Survivorship issues following treatment completion-results from focus groups with Australian cancer survivors and health professionals. J Cancer Surviv. 2008;2(1):20–32.

[5] McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin. 2011;61(1):50–62.

[6] Barlow JH, Bancroft GV, Turner AP. Self-management training for people with chronic disease: a shared learning experience. J Health Psychol. 2005;10(6):863–72.

[7] Davies NJ, Batehup L. Self-management Support for Cancer Survivors: Guidance for Developing Interventions. An Update of the Evidence (2010) Self-Management Work stream NCSI/Macmillan Cancer Support. 2010.

[8] Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns. 2002;48(2):177–87.

[9] Cockle-Hearne J, Faithfull S. Self-management for men surviving prostate cancer: a review of behavioural and psychosocial interventions to understand what strategies can work, for whom and in what circumstances. Psycho-oncology. 2010;19(9):909–22.

[10] Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res. 2013;15(2):e37.

[11] Greene SM, Tuzzio L, Cherkin D. A framework for making patient-centered care front and center. Perm J. 2012;16(3):49–53.

[12] Dickerson SS, Boehmke M, Ogle C, Brown JK. Seeking and managing hope: patients' experiences using the Internet for cancer care. Oncol Nurs Forum. 2006;33(1):E8–17.

[13] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

[14] Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res. 2010;12(1):e4.

[15] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration; 2011 [accessed November 15, 2014]. Available from: http://www.cochrane.org/hand-book.

[16] Hedges LV, Olkin I. Estimation of a single effect size: parametric and nonparametric method. In: Statistical Methods for Meta-Analysis. San Diego, Academic Press, 1985, pp. 75–106.

[17] Wilson DB, Lipsey M. Practical Meta-analysis. Thousand Oaks, CA: Sage; 2001, pp. 146–8.

[18] Galnz K, Rimer, B.K., Biswanath, K. Health Behavior and Health Education: Theory, Research, and Practice. San Fransisco, CA: Jossey-Bass; 2008.

[19] Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med. 2010;51(3–4):214–21.

[20] Yun YH, Lee KS, Kim YW, Park SY, Lee ES, Noh DY, et al. Web-based tailored education program for disease-free cancer survivors with cancer-related fatigue: a randomized controlled trial. J Clin Oncol. 2012;30(12): 1296–303.

[21] Hammer MJ, Ercolano EA, Wright F, Dickson VV, Chyun D, Melkus GD. Self-management for adult patients with cancer: an integrative review. Cancer Nurs. 2015; 38(2):E10–26.

[22] Finfgeld DL. Therapeutic groups online: the good, the bad, and the unknown. Issues Ment Health Nurs. 2000;21(3):241–55.

[23] Jang EH, Choi KS, Kwon SJ. A meta-analysis of intervention on depression in cancer patients. Asian Oncol Nurs. 2013;13(1):1–10.

Address for correspondence

Hyeoun-Ae Park, PhD

College of Nursing and Systems Biomedical Informatics Research Center, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea. Tel.: +82-2-7408827, Fax: +82-2-7654103, E-mail: hapark@snu.ac.kr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-148

The Shared Decision Making Frontier: a Feasibility and Usability Study for Managing Non-Critical Chronic Illness by Combining Behavioural & Decision Theory with Online Technology

Amina Russell[a], William Van Woensel[a], Samina Raza Abidi[b]

[a] *Department of Computer Science, Dalhousie University, Halifax, NS, Canada*
[b] *Department of Medicine, Dalhousie University, Halifax, NS, Canada*

Abstract

The objective of this study is to determine if shared decisions for managing non-critical chronic illness, made through an online biomedical technology intervention, us feasible and usable. The technology intervention incorporates behavioural and decision theories to increase patient engagement, and ultimately long term adherence to health behaviour change. We devised the iheart web intervention as a "proof of concept" in five phases. The implementation incorporates the Vaadin web application framework, Drools, EclipseLink and a MySQL database. Two-thirds of the study participants favoured the technology intervention, based on Likert-scale questions from a post-study questionnaire. Qualitative analysis of think aloud feedback, video screen captures and open-ended questions from the post-study questionnaire uncovered six main areas or themes for improvement. We conclude that online shared decisions for managing a non-critical chronic illness are feasible and usable through the iheart web intervention.

Keywords:

Shared Decision Making; Behavioural Medicine; Choice Architecture; Biomedical Technology Intervention; Chronic Illness.

Introduction

Shared decision making (SDM) is considered the cornerstone of patient-centred care [1], signifying an important paradigm shift away from paternalistic medicine. In SDM, the patient and physician collaborate to select the best diagnostic and treatment options. It is a meeting of experts, in which the physician is the expert in medicine and the patient is the expert in his or her own life, values and circumstances [2].

Clinical trials have demonstrated that SDM can improve adherence to treatment and clinical outcomes [3, 4]. Despite its proven effectiveness, literature shows that only 10% of face-to-face clinical consultations involve SDM [5]. Biomedical technology interventions, also known as decision aids, have sought to fill this SDM gap [6]. Web-based apps aimed at improving lifestyles (i.e., weight change, nutrition, and physical activity) show evidence of positive impacts [7]. One study [8] concludes that web-based interventions increase patient activation and have the potential to enhance the self-management capabilities of the growing population of chronically ill people. A systematic review of internet-based interventions for hypertension revealed they significantly reduced systolic blood pressure by 3.8 mm Hg and diastolic blood pressure by 2.1 mm Hg [9]. In the Netherlands where heavy drinking among young adults has become a public concern, a tailored web-based intervention aims to reduce drinking practices amongst college students [10].

Use of behavioural theories improves the success rate of these health interventions [11]. For example, theory-driven behavioural strategies can reduce the risk of cardiovascular disease through lifestyle change rather than increasing medication [12]. Many different behavioural theories exist such as the Integrated Change (I-Change) behavioural model which has three simple states: awareness, motivation and action [13]. Its straightforwardness makes it a good fit for biomedical technology interventions, which should be kept simple as they involve patient interaction.

Choice architecture (CA) is another technique for improving decisions, as well as their implied long-term behaviour change (e.g., reduce smoking) [14]. CA is a decision theory originating from the field of economics [15] and has been successfully used to improve adherence to medication use [14]. It serves to fill the well-known intent-behaviour gap [16], characterized by a disconnection between intention and actual behaviour.

Reinforcement through multiple methods may increase patient motivation, and adherence to decided-upon health changes [17]. Both behavioural and choice theories are strategies aimed at improving patient engagement, which is an essential ingredient to SDM. The objective of this study is to demonstrate that behavioural theory and decision theory can complement one another within a technology intervention for SDM.

Methods

The iheart web application is a "proof of concept" to determine whether SDM for the management of a non-critical chronic illness, using a biomedical technology intervention, is feasible and usable. In the software development lifecycle, feasibility determines whether it is sensible to develop a system by objectively reviewing its strengths, weaknesses and opportunities, for instance via proof-of-concept functional prototypes [18]. Usability according to the ISO 9241 standard is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [19].

The iheart solution evolved over five phases: 1) conceptual, 2) design, 3) application development, 4) testing and 5) assessment of feasibility and usability.

Conceptual Phase

The initial blueprint for the biomedical technology intervention, employing the I-Change behavioural theory and CA, was first defined in a conceptual process flow diagram. The diagram visually outlined the patient navigation when using the system, indicating the different choices and SDM steps between patient and provider.

Design Phase

This conceptual diagram informed the application design. Firstly, the employed behavioral theories, SDM steps, and domain-specific content (i.e., related to hypertension) were represented as an ontology knowledge model. The logic rules for the application, including the SDM logic, were documented externally in decision tables, along with domain-specific messaging.

For hypertension, behaviour change choices include increasing exercise, reducing smoking or reducing sodium. Depending on their current lifestyle, the patient is presented with one or more of these options, and chooses one based on SDM with the healthcare provider. Afterwards, the patient fills out a behavioural questionnaire, which scores his/her initial I-Change behavioural state. I-Change is a behavioural theory with three behavioural states (awareness, motivation and action) and determines the patient's readiness for making the selected behaviour change. Importantly, this state informs elements of the CA, a decision theory that attempts to fill the intent-behaviour gap. In particular, based on the patient's behaviour state, appropriate CA messaging was presented, adapted to his/her readiness for making the selected behaviour change. An expert in Industrial-Organizational Psychology devised and validated the behavioural questionnaire. Questionnaires, content and informational messages were further reviewed by a nurse to ensure they were patient appropriate and also met patient education readability guidelines [20]. Finally, the patient was able to set a concrete goal given his/her health, again during an online SDM chat session with the healthcare provider.

The ontology model was built to be flexible, accommodating any disease condition, the selection of different behavior change theories and different CA concepts; as well as scalable, to support multiple health care providers and patients. This patient-centric model further contained the core constructs, relationships, content, messages and description logic that formed the foundation for the technical solution.

We devised mock screens based on the ontology model, rules and conceptual flow. An entity-relationship diagram defined the underlying database model for the application.

Application Development

The biomedical technology solution consisted of a web-based interface storing information on a centralized server within a secure relational database (Figure 1). The Vaadin open source web application framework [21] was used to produce an internet application accessible on PCs and mobile devices. Vaadin has a plug-in for chat sessions that enabled online SDM.

The iheart data was physically stored in a MySQL database [22] on a secure server. Eclipselink [23] was used to automatically persist application objects in the database, and implement the Java Persistence API (JSR 317) [24]. The Drools business management system [25] implements the Java Rules Engine API (JSR 94) [26], and was used for implementing decision logic based on decision tables. This decision logic included suggesting behaviour changes, based on the patient's current lifestyle; and determining CA messages depending on the patient's I-Change behavioural state. Finally, the Big Life Sodium Calculator [27], a third party web service, was made accessible from iheart, and calculated the participant's daily sodium consumption based on a series of questions.

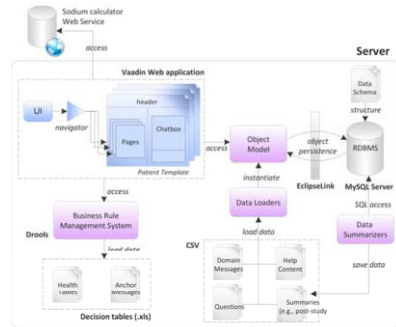


Figure 1 – iheart application architecture

Testing Phase

SDM via an online chat system (Figure 2) represents a novel approach for healthcare. Thus, several rounds of testing were applied to objectively prepare the system for evaluation. First, two medical doctors and two hypertensive patients critically reviewed the first-cut of the iheart prototype. Adjustments followed in response to this initial, informal feedback.

Figure 2 – iheart decision making and chat box

The formal application testing involved both white box and black box testing. An inward look at the application (white box) focused on unit and integration testing of internal components. An outward look at the application (black box) tested its external and user facing features. For example, the black box portion validated the seamless use of the Big Life Sodium Calculator web service within iheart. During both types of testing, modifications were made in iterations followed by regression testing of the application.

Study Assessment

The assessment of iheart was completed through a pilot study approved by the research ethics board of Dalhousie University, Canada. Only four to five subjects are needed to identify 80% of the usability problems with a system [28]. The study recruited nine hypertensive participants, approximately twice the recommended number, from the local area for this purpose.

Three goal-based scenarios were devised in collaboration with a medical doctor to assess the feasibility and usability of iheart. The scenarios were intentionally devised to present more than one lifestyle choice for reducing hypertension. This would compel the participants to make shared decisions. Each participant completed two of the three scenarios in order to sufficiently assess and rate the iheart application. Each scenario was used six times by the pool of nine participants.

This ensured uniformity and consistency in the assessment of the application.

Results

Quantitative Analysis

A post-study questionnaire solicited the participant's usability scores using a five-point Likert scale, combined with open ended questions for capturing experience and thus allowing qualitative analyses (see next section). The questionnaire contained sections on the usability, content, CA, SDM and overall functionality of iheart. The Likert Scale used: 1=Strongly Agree, 2=Moderately Agree, 3=Neither Agree nor Disagree, 4=Moderately Disagree and 5=Strongly Disagree. A score of 1 or 2 favoured iheart whereas scores of four or five indicated discontentment.

The questionnaire contained 41 Likert scale questions. Statistical software calculated the mean, median and mode scores within each section of the post-study survey (Table 1). Further analysis transpired at the participant level (Table 2).

Table 1 – Section Level Data

Section	Mean	Median	Mode	Chronbach's Alpha
Usability	2.07	1.5	1	0.97
Content	2.35	2.0	1	0.94
Choice	1.98	1.0	1	0.85
SDM	2.09	1.0	1	0.95
Overall	2.13	1.0	1	0.97

Within survey sections in Table 1 showed participant satisfaction across the board. Chronbach's alpha validated response consistency within survey sections.

Table 2 – Participant Level Data

Participant	Mean	Median	Mode
p1	1.44	1.0	1
p2	1.34	1.0	1
p3	1.80	1.9	1
p4	1.05	1.0	1
p5	1.61	1.0	1
p6	3.49	4.0	4
p7	4.24	4.0	5
p8	2.95	3.0	2
p9	1.24	1.0	1

Table 2 demonstrated that six participants (two-thirds) favoured the iheart application while participant 8 appeared neutral, participant six leaned towards dissatisfaction and participant seven appears dissatisfied. This quantitative data, when collated with the qualitative information, identified specific areas for improvement to iheart.

Qualitative Analysis

The study captured qualitative data primarily using the "think aloud protocol" [29]. This protocol captures participant feedback spoken out loud while using iheart to achieve the goal-based scenarios. The software, Active Presenter [30], captured the audio and video feedback including screen captures. Participants also responded to three open ended questions in the post-study questionnaire.

The qualitative data was imported into a qualitative analysis tool, ATLAS.Ti [31]. Inductive thematic coding [32] occurred in two stages: 1) open coding and 2) axial coding [33]. First, a comprehensive code list was increasingly built as each piece of qualitative data was openly coded. Open coding involves reading through the data several times, and labeling chunks of data. Afterwards, the frequency or "groundedness" of each code (Figure 3) was reviewed to identify the most common feedback, either positive or negative (prefixed with a plus and minus sign, respectively).

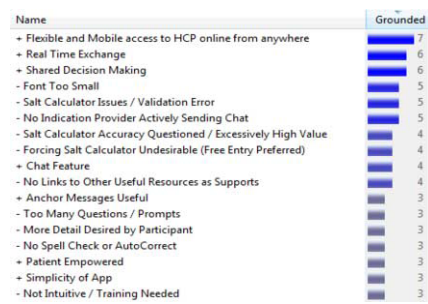


Figure 3 – Participant Feedback by Frequency

Axial coding was then applied to draw categories from the open code list. Axial coding identifies commonalities or relationships amongst the open codes. The "family manager" in the ATLAS.Ti software managed these categories or themes. The derived themes and their frequencies were: communication (10), usability (8), content (8), user interface (9), features (7), chat (6), salt calculator (3) and a miscellaneous (3) category. The themes included both positive and negative feedback, meaning each theme represents areas for improvement. Sorting by theme gives a detailed account of the specific improvements needed. Analysis of the negative feedback showed that all desired improvements were feasible, including functional (e.g., no indication a chat is being sent back), informational (e.g., links to other resources desired) or cosmetic (e.g., font size too small) in nature.

Discussion

A mixed methods study provided rich feedback on possible refinements for the iheart web application. In particular, the qualitative data presented a wealth of information on the benefits and challenges of the intervention that was not apparent in the quantitative data. For instance, the usability section of the post-study questionnaire had a mean score of 2.07, implying the participants moderately agreed the application was usable. Nonetheless, six items were identified in the qualitative analyses to improve the user interface and four items to improve the chat feature. Both the user interface and chat feature contribute to usability.

Additionally, the qualitative feedback identified other practical situations where iheart could be applied to manage chronic illness. In particular, participants felt that iheart would

be very useful in general long term care where a physician is not always present onsite. They also saw value in situations where patient mobility is limited (e.g., cannot leave home).

Vision

A notable feature of iheart is its scalability. The application was designed with expansion in mind. The ontology knowledge model accommodates decision-making with more than one health provider per patient to accommodate multi-disciplinary teams of medical professionals providing care. Due to its reusable components, the web application can be customized for other chronic illnesses as well, thus expanding its use to other medical domains.

The study data demonstrates that participants most appreciated the flexibility, mobility, shared decisions and real time chat exchanges with a remote healthcare provider. This suggests a paradigm shift within the practice of traditional medicine. It means medical providers, endorsing a tool such as iheart, would have to accommodate scheduled "chat time" in their daily or weekly schedule to interact online in a real-time fashion with patients. Or, similar to an online call centre, the configuration could include recruitment of a pool of qualified medical professionals to provide real-time interaction with patients. It is the next frontier in online medicine. Patients are increasingly seeking timely and informative medical answers online. Currently, they seek it in the form of static information (e.g., webMD knowledge bases). Based on our study, it seems that online, human interactive exchange is another appealing knowledge medium for patients.

Limitations

A limiting element of iheart was the loss of face-to-face communication currently used in traditional medicine. Some participants expressed that without facial expressions, body language and visual cues, it was difficult to read the healthcare provider and also describe their medical situation adequately. Seeing how patients appreciated the flexibility and mobility of the real-time text-based chats, this challenge could be addressed through the use of a video chat option. Patients could alternatively have a combined video and text chat (much like a Snapchat session) during shared decision points to reduce the communication void from a text-only session.

Future Work

The iheart application should be revised based on the study feedback. Moreover, a logical next step would be performing an efficacy study, to assess adherence to the application's daily use for managing chronic illness such as hypertension. This study should also solicit input and feedback from a larger selection of healthcare providers.

To further assist in decision making, behaviour change choices available during SDM could be accompanied by evidence-based metrics. A health informatics approach could capture each participant's intervention choice and their adherence to the particular behaviour. After a fixed number of participants (e.g., 100) have used iheart, the cumulative data from people with similar demographics could be used to indicate success rates for each behaviour choice.

Conclusion

The majority of study participants positively perceived the iheart application. Think aloud and qualitative feedback identified specific areas for improvement. The desired improvements are all possible with the exception of the salt calculator, which needs to be changed by its creators. Overall, making shared decisions through a biomedical technology intervention proved both feasible and usable.

As Canada's population is set to enter a period of relatively rapid aging [34], we need to pursue innovative means of delivering care. The Canadian trend for elderly patients with chronic illness or low mobility is to seek care in their homes [35]. As indicated by our test subjects, a system like iheart could be a good fit for residential and long term care of non-critical chronic illness. We recommend the revised version of iheart be pilot tested in a variety of settings where patients have low mobility or challenges reaching their physicians.

Acknowledgments

This study was supported by a NICHE Research grant under the oversight of Dr. Syed Sibte Raza Abidi. The Big Life Salt Calculator was devised by the Ottawa Hospital Research Institute and Dr. Douglas Manuel. Dr. Nelofar Kureshi, Dr. Susan Alexander, Cheryl Merkel and Ann Dent provided feedback or domain expertise to this study.

References

[1] Barry MJ, Edgman-Levitan S. Shared decision making - the pinnacle of patient centred care. *N Engl J Med* 2012; 366(): 780-81. http://www.nejm.org/doi/full/10.1056/NEJMp1109283 (accessed 18 October 2014).

[2] Tuckett D, Boulton M, Olson C, Williams A. Meeting between experts: An approach to sharing ideas in medical consultations. New York: Routledge; 1986.

[3] Wilson SR, Strub P, Buist SA, Knowles SB, Lavori PW, Lapidus J, Vollmer WM. Shared Treatment Decision Making Improves Adherence and Outcomes in Poorly Controlled Asthmas. *Am J Respir Crit Care Med* 2010; 181(6): 566-77.

[4] Mandelblatt J, Kreling B, Figeuriedo M, Feng S. What Is the Impact of Shared Decision Making on Treatment and Outcomes for Older Women With Breast Cancer?. *J Clin Oncol* 2006; 24(30): 4908-13. http://jco.ascopubs.org/content/24/30/4908.full.pdf (accessed 21 November 2014).

[5] Godolphin W. Shared decision making. *Healthc Q* 2009; 12(Special Issue): 186-90. http://www.ncbi.nlm.nih.gov/pubmed/19667767 (accessed 18 October 2014).

[6] Ozanne EM, Annis C, Adduci K, Showstack J, Esserman L. Pilot trial of a computerized decision aid for breast cancer prevention. *Breast J* 2007; 13(3): 147-54. http://www.ncbi.nlm.nih.gov/pubmed/17319855 (accessed 18 October 2014).

[7] Safran Naimark J, Madar Z, R Shahar D. The Impact of a Web-Based App (eBalance) in Promoting Healthy Lifestyles: Randomized Controlled Trial. J Med Internet Res. 2015;17(3):e56.

[8] Solomon M, Wagner S, Goes J. Effects of a Web-Based Intervention for Adults With Chronic Conditions on Patient Activation: Online Randomized Controlled Trial. J Med Internet Res. 2012;14(1):e32.

[9] Liu S, Dunford S, Leung Y, Brooks D, Thomas S, Eysenbach G et al. Reducing Blood Pressure With Internet-Based Interventions: A Meta-analysis. Canadian Journal of Cardiology. 2013;29(5):613-621.

[10] Voogt C, Poelen E, Kleinjan M, Lemmers L, Engels R. The development of a web-based brief alcohol intervention in reducing heavy drinking among college

students: an Intervention Mapping approach. Health Promotion International. 2013;29(4):669-679.

[11] Office of Behavioral & Social Sciences Research. *Social and Behavioral Theories.* http://www.esourceresearch.org/eSourceBook/SocialandBehavioralTheories/1LearningObjectives/tabid/724/Default.aspx (accessed 18 October 2014).

[12] Cox JL, Vallis TM, Pfammatter A, Szpilfogel C, Carr B, O'Neill. A Novel Approach to Cardiovascular Health by Optimizing Risk Management (ANCHOR): behavioural modification in primary care effectively reduces global risk. *Can J Cardiol* 2013; 29(11): 1400-07. http://www.ncbi.nlm.nih.gov/pubmed/23796526 (accessed 18 October 2014).

[13] de Vries H, Mesters I, Van de Steeg H, Honing C. The general public's information needs and perceptions regarding hereditary cancer: an application of the integrated change model. *Patient Educ Couns* 2005; 56(2): 154-56. http://www.ncbi.nlm.nih.gov/pubmed/15653244 (accessed 18 October 2014).

[14] Nease RF, Frazee SG, Zarin L, Miller SB. Choice architecture is a better strategy than engaging patients to spur behavior change. *Health Aff (Millwood)* 2013; 32(2): 242-49. http://www.ncbi.nlm.nih.gov/pubmed/23381516 (accessed 18 October 2014).

[15] Schmidt A. *Choice Architecture: Helping patients make better health decisions by design.* http://www.anderson.ucla.edu/assets-digital/think-in-the-next/choice-architecture-dept (accessed 18 October 2014).

[16] Fennis BM, Adriaanse MA, Stroebe W, Pol B. Bridging the intention–behavior gap: Inducing implementation intentions through persuasive appeals. *J Consum Psychol* 2011; 21(3): 302-11. http://www.sciencedirect.com/science/article/pii/S1057740810001439 (accessed 18 October 2014).

[17] Leddy MA, Anderson BL, Schulkin J. Cognitive-behavioral therapy and decision science. *New Ideas Psychol* 2013; 31(3): 173-83. http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.elsevier-d837f335-13c7-3d8a-bb97-6794f77cb9b4 (accessed 18 October 2014).

[18] Dictionary.com. *Feasibility study.* http://dictionary.reference.com/browse/feasibility+study (accessed 22 November 2014).

[19] Usability Professionals Association. *What is usability?.* http://www.usabilityprofessionals.org/usability_resources/about_usability/definitions_of_usability.html (accessed 22 November 2014).

[20] Capital Health. *Patient and Family Education Material Guidelines.* http://www.cdha.nshealth.ca/health-sciences-library/documents (accessed 23 October 2014).

[21] Vaadin 7.3.3. [Computer Software]. Turku Finland.

[22] Oracle Corporation, MySQL. [Computer Software]. Retrieved from http://www.mysql.com/

[23] The Eclipse Foundation, EclipseLink. [Computer Software]. Retrieved from http://www.eclipse.org/eclipselink/

[24] EJB 3.0, Java Persistence API. [Computer Software]. Retrieved from http://www.oracle.com/technetwork/java/javaee/tech/persistence-jsp-140049.html

[25] Red Hat, Inc., Drools. [Computer Software]. Retrieved from http://www.drools.org/

[26] Oracle Corporation, JSR-000094 Java™ Rule Engine API. Retrieved from https://www.jcp.org/aboutJava/communityprocess/final/jsr094/

[27] Manuel D, L'Abbe M, Arcand J. Big Life Sodium Calculator [Internet Service]. Retrieved from http://www.projectbiglife.ca/sodium/ (accessed 23 October 2014).

[28] Vizri RA. Refining the test phase of usability evaluation: how many subjects is enough?. *The Journal of the Human Factors and Ergonomics Society* 1992; 34(4): 457-68.

[29] Benbunan-Fich F. Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management* 2001; 39(2): 151-163. http://www.sciencedirect.com/science/article/pii/S0378720601000854 (accessed 25 October 2014).

[30] Atomi Systems Inc., Active Presenter. [Computer Software]. Retrieved from http://atomisystems.com/activepresenter/

[31] Scientific Software Development GmbH, Atlas.ti. [Computer Software]. Retrived from http://atlasti.com/

[32] Boyatzis RE. *Transforming Qualitative Information*. Thousand Oaks, California: SAGE Publications, Inc.; 1998.

[33] Pandit NR. The creation of theory: A recent application of the grounded theory method. *The Qualitative Report* 1996; 2(4):. http://www.nova.edu/ssss/QR/QR2-4/pandit.html/pandit.html (accessed 07 November 2014).

[34] Canadian Institute for Health Information. Healthcare in Canada 2011: A focus on seniors and aging. 2011; (): . https://secure.cihi.ca/free_products/HCIC_2011_seniors_report_en.pdf (accessed 08 November 2014).

[35] Statistics Canada. *Type of housing of a primary care receiver differs by age and sex.* http://www.statcan.gc.ca/pub/11-008-x/2008002/t/10689/5801409-eng.htm (accessed 08 November 2014).

Address for correspondence

Amina Russell, Dalhousie University

E-mail: arussell@cs.dal.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-153

Experiences of Healthcare Professionals to the Introduction in Sweden of a Public eHealth Service: Patients' Online Access to their Electronic Health Records

Ture Ålander[a], Isabella Scandurra[b]

[a] *Department of Public Health and Caring Sciences, Uppsala University, Sweden*
[b] *School of Business, eHealth, Örebro University, Sweden*

Abstract

Patients' increasing demands for medical information, the digitization of health records and the fast spread of Internet access form a basis of introducing new eHealth services. An international trend is to provide access for patients to health information of various kind. In Sweden, access by patients to their proper electronic health record (EHR) has been provided in a pilot county since November 2012. This eHealth service is controversial and criticism has arised from the clinical professions, mainly physicians. Two web surveys were conducted to discover whether the opinions of healthcare professionals differ; between staff that have had experience with patients accessing their own EHR and those who have no such expericence. Experienced nurses found the EHR more important for the patients and a better reform, compared to unexperienced nurses in the rest of the country. Similarly, physicians with their own experience had a more positive attitude compared to non-experienced physicians. The conclusion of this study is that healthcare professionals must be involved in the implementation of public eHealth services such as EHRs and that real experiences of the professionals should be better disseminated to their inexperienced peers.

Keywords:

Electronic Health Records; Health Personnel; Online Systems; Professional-Patient Relations; Doctor-Patient Relations; eHealth; Access to Information; Healthcare Survey; Patient Participation

Introduction

The increasing demands of patients to read medical information pertaining to themselves, the digitization of health records and the fast spread of Internet access form the foundation of the implementation of a service giving patients direct online access to their electronic health records (EHR). All Swedish patients have a right to obtain a part of their health records [1] and the national eHealth strategy states that all health records should be accessible via Internet (i.e., online electronic health records) by 2017 [2].

Also in European directives, such a solution is sought to gain increased patient safety and security [3]. As the introduction of access to health records by patients and their families is an international trend, findings from research on pilot sites with real experience are of wide interest.

However, patients' access to their own health record is a controversial issue with many aspects. It is since a long time stated that the computerization transforms the accountability of professionals in relation to different actor groups: patients as well as managers, politicians and auditors [4]. The autonomy of the medical profession in relation to organisational governing and control systems has been discussed [5]. A common fear among the professionals is that the autonomy of the medical profession is expected to change and decrease. Less attention has so far been drawn towards professional transparency versus autonomy in relation to the patients, for which this kind of studies provides a basis. With this study, we want to focus on the impact on professionals due to patients new EHR transparency.

In Sweden, EHR systems were adopted quite early. Internet access as well as mobile phones have been widespreaded in the population from the early 1990's. In 1997, Uppsala County Council in Sweden started a project with the aim to give patients access to their medical data. The project called Sustains had financial support from the European Commission [6]. A system was developed to allow patients to have direct access to their clinical notes as well as to several other eServices through a "Healthcare account" similar to a bank account, over the Internet [6]. The system was introduced and used on a solo family practice in Uppsala and tested in clinical practice for many years [7].

The fast digital development and experiences from the Sustains project [6,7] pushed forward a change of the Swedish legislation in 2008 [1], which permitted the healthcare organizations to give patients direct access to their EHRs including the laboratory values and doctors' notes.

All hospitals and primary healthcare centers, except community nurses and some private doctors, in Uppsala County Council, used the same system developed by Cosmic, from a Swedish healthcare IT company: Cambio Healthcare Systems.This facilitated the initial integration work between the region-wide EHR and the public eHealth services.

Within the scope of an EU-project [8] Uppsala County Council in august 2012 introduced a new eHealth service called "My medical record" with the content from the Cosmic EHR. With this eHealth service, the patients from the age of 18 years old have access to all written text, including doctors' notes, on the Internet. The system was tested with employed health personnel in Uppsala County Council before it was launched [9].

In November 11th 2012 the e-service was offered to all 350,000 registered inhabitants in Uppsala County Council and soon later, in December 3rd, to all Swedish inhabitants with EHRs in the Uppsala County. As the first region-wide trial, the patients were given access to their health records through a secure log in to the national eHealth patient portal "My Healthcare Contacts" (www.minavardkontakter.se). People log in to the portal using the same general electronic ID as they use for banking and government e-services. Other e-

services successively introduced in the pilot county through the national patient portal were to:

- book, re-book/re-schedule or cancel an appointment,
- request certificates,
- extend sick leave,
- create safe messages with nurse or doctor,
- update personal data,
- change house physician/family doctor,
- renew medical prescriptions and assistive tools,
- and to order a written copy of the proper medical health record [10].

Apart from reading the medical text, the patient also has access to the audit log, block access to their records, give access to a friend or relative, order, get a SMS reminder of appointments, get a written list of prescriptions and laboratory values, can fill in a health declaration form saved in their records, and see the status and entire flow of their referrals to specialists or hospital. Both authors of this paper have access to their own health record and current version of views of data starts with either a time line or a calendar view, of the patient's choice.

From the release of the eHealth service in the pilot county, the number of users was growing with about 100 new users every day and in September 2013 6.4% of the eligible patients (n=477,928) were users [10]. By the end of 2014 the service had 48,664 unique users that could access their own and/or their family members' records [10]. In March 2015 there are 80,000 users who have had an average of four accesses to their health record and 800 patients accessed their health records every day [11,12].

Although the eHealth service is popular among patients, it is controversial among many professionals, especially physicians [13]. It is evident that increased knowledge of how the e-service influence users, both patients and clinicians, is essential for successful deployment of public eHealth [14] and that such services in general are challenging to put into practice.

This was the first region-wide trial in Sweden. The introduction of patients' access to their EHR creates an unique possibility to examine both longitudinal changes as well as momentary aspects related to professional transparency, patient relations, and patient empowerment.

The study presented in this article is part of a larger research project, DOME, Deployment of Online Medical Records and eHealth services in Sweden [15] with the aim to highlight experiences and effects related to the introduction of public eHealth services. The project was created in July 2012 in order to connect the first European deployment project, Sustains [8], to a purposive research group consisting of 16 nationally spread researchers from various scientific fields. This multi-site and multi-disciplinary composition provides a unique opportunity to highlight the issues from various research aspects through different methods and studies. Currently the senior researchers cover the areas of information management, human-computer interaction, IT and work environment, management and business studies, information security, healthcare informatics, medicine, organization theory, eGovernment, information technology, engineering education and statistics [15].

Different evaluation studies are taking place within three work packages, where this work is part of comparisons of surveys directed towards health care professionals in work package B, professions and management. The research questions for this study were:

- Which were the attitudes of the nurses and the physicians after they had been working some time within the pilot county compared to professionals that were not yet experienced with the eHealth service?
- Do experiences from the pilot county differ from regions where patient access to Electronic Health Records are not yet implemented?

Materials and Methods

Patients from Uppsala County Council who were treated in hospital and primary care and patients from other counties who had received care in Uppsala County Council all have direct access to their EHRs since December 2012. The service is optional to use for patients and there is no requirement for staff or patients to talk about the usage of the service.

The two web surveys of this study were distributed to the physicians in Uppsala county 6 months after the deployment of the new e-service and to the nurses 9 months after the deployment. Data were collected from two similar 5-graded Likert scale web surveys (S1 and S2) to Swedish healthcare staff and is focused on attitudes and opinions of physicians (S1) and nurses (S2). To each set of statements, there was also a possibility to give answers in free text.

S1 was sent out in June 2013 to 1,602 physicians in the pilot county with a response rate of 25% (399 respondents, 52% women, age 25-71 years, median worktime 14 years, 78% worked inhospital and 22% in primary care, 85% native in Swedish language).

S2 was sent out in March 2014 to 8,460 registered nurses and midwives in Sweden with a response rate of 35,4% (2,867 respondents, of whom working as nurses were 84%, midwives 6%, chief position 5%, in projects 2% and other 3%). The questionnaires consisted of background questions and five sets of items with free text fields to each set. To deliver the questionnaires the Uppsala County Council's and the Swedish Association of Health Professionals' internal web survey tools were used. The accompanying letter stated that responding was voluntary and that the time spent to respond according to the strongly agree (5) - strongly disagree (1) - scale was approximately 10 minutes.

Ethical approval for S1 was granted by the Uppsala County Council's research units as well by the Swedish central ethical committee. S2 was conducted according to the principles of the Declaration of Helsinki.

Standard data reports were created for each survey with charts showing the most prominent differences in each statement. As the questionnaires were jointly developed and contained some identical statements, this first analysis expects to discover differences between nurses and physicians that have experience of patients accessing their EHRs online and those working in regions where the eHealth service had not yet been implemented. The two different questionnaires distributed to the nurses and physicians contain more statements to analyze, however the responses handled in this study are consistent between the two surveys. Currently other statements from the data-sets are jointly being analyzed by statistics, clinical and healthcare informatics researchers and students and will be published in the near future.

The statements analysed in this study regard to which extent patients' access to EHRs online is considered useful for the patients, if their relatives were given access to their EHRs and whether the professionals consider the new eHealth service being a good reform. In the questionnaires the statements read:

- To which extent do you consider it a benefit for the patients that their relatives have a possibility to take part in their EHR?

- To which extent do you consider the eHealth service "online health records" a good reform?

Simple statistics were used for analysis. The Mann-Whitney ranksum and χ^2 tests were used for evaluating the statistical significance for ordinal and nominal data, respectively. The data was analyzed by the Stata statistical package 13.1 [16].

Results

The responses of the physicians to the multiple choice questions are presented in Table 1 based on a 5-graded scale from strongly disagree (1) to strongly agree (5). The responses of the nurses are presented in Table 2 and Table 3. In general, professionals' opinions of patients reading their health record online were negative (grade<2.5) among physicians and positive (grade>2.5) among nurses. Physicians were negative to giving relatives of the patients access to the EHR and they were also negative to the reform as such. Both pilot nurses and those outside the pilot county were generally more positive than the physicians.

Nurses in Uppsala County Council with experience with EHRs were significantly more positive to reform ($p<0.001$) than nurses outside the pilot county. 40% stated that they agreed or strongly agreed to 'patients receiving online access to their EHRs being a good reform' compared to nurses in the rest of the country (25%) as well as physicians, where 82% disagreed or strongly disagreed.

Table 1 – Physicians working in the pilot county (Uppsala County Council) Responses to the statement "To which extent is it a benefit for the patients that their relatives have a possibility to take part in their EHR?" and whether the eHealth service "is a good reform".

Five-graded Likert scale 1-5 where 5 = strongly agree.

	Access for relatives	A good reform
N	381	385
Mean	2.22	1.66
SD	1.08	0.97

In one of the background questions of the web questionnaire, 61 physicians of 399 (15%) stated that they for themselves or their relatives had personal experience of using this eHealth service as patients.

A sub-analysis revealed that their own usage was equal between hospital and primary care physicians, females and males, but that the users were younger ($p=0.0004$).

Physicians using the eHealth service themselves, compared to non-users, regarded the eHealth service:

- more important for patients ($p=0.0001$),
- to improve quality in care ($p<0.0001$),
- to contribute more to equality in healthcare ($p=0.0003$),
- they experienced less change in their working conditions ($p=0.014$),
- found the eHealth service being a good reform to a greater extent ($p=0.003$),
- thought their medical noting had improved since the eHealth service was introduced ($p=0.021$) and
- found the launch of the reform better ($p=0.019$).

Table 2 – Nurses working in and outside the pilot county (Uppsala County Council). Responses to the statement "To which extent do you consider it a benefit for the patients that their relatives have a possibility to take part in their EHR?"

Five-graded Likert scale 1-5 where 5 = strongly agree.

	Pilot county	Outside Pilot	P (Mann-Whitney)
N	234	2495	
Mean	3.08	2.92	0.0454
SD	1.16	1.12	

Table 3 – Nurses working in and outside the pilot county response to the statement "To which extent do you consider the eHealth service "online health records" a good reform?"

Five-graded Likert scale 1-5 where 5 strongly agree.

	Pilot county	Outside Pilot	P (Mann-Whitney)
N	235	2494	
Mean	3.17	2.71	<0.0001
SD	1.27	1.17	

Again, nurses from the pilot county were far more positive (35% agree and strongly agree) to the importance of this eHealth service for the patients, compared to nurses outside the pilot (25%) as well as physicians (64% disagree or strongly disagree).

Discussion

This study aimed to investigate the opinions of health professionals related to the introduction of a new eHealth service: giving patients direct access to their EHRs, in Sweden. The main outcome of this study was that many professionals were reluctant with the proposed eHealth service, and that professionals with personal experience with the innovation were more positive as to the benefit for patients than healthcare professionals without experience.

Nurses with experience with the service were significantly more positive to reform ($p<0.0101$) than nurses outside the pilot region. The physicians within the pilot region were generally more negative compared to nurses in the same region. However, the time spent between the questionnaires (9 months) does cause an uncertainty as to comparability between the opinions of the professionals. The longer use time for nurses compared to physicians may have given the nurses a more positive attitude.

The authors' knowledge about the context in which this deployment process took place is worth mentioning. During the deployment project [8], representatives of the physicians' local union in the pilot county expressed a distrust of the eHealth service, whereas the Swedish Association of Health Professionals (the union for the nurses and midwives) embraced the development and deployment of such eHealth services. These different statements could have affected the respondents of this study.

However, in another study, free text responses from the survey of the physicians were analyzed and revealed that the physicians in general were negative and many also upset over the deployment of an eHealth service giving all patients digital access to their medical records. The physicians were especially concerned about increased workload and patients' risk of misunderstanding the notes and laboratory findings. These findings are further presented by Erlingsdottir et al [13].

These findings correspond to the hypothesis that physicians have a wider knowledge and a stronger patient relationship

than nurses. Thus they feel a greater threat to their autonomy and thus get a more negative attitude to online EHRs for patients.

As the patients become more knowledgeable, the privileged position of the professionals changes. As early as in 1985, Freidson argued that computerization may contribute to the decreasing of the knowledge gap as more and more information becomes easily accessible [4]. According to Freidson, the key question is whether laymen are able to mobilize sufficient motivation to become knowledgeable to the extent that they will be able to challenge professional actors. Possibly, the resistance that is noted in this study could be related to the intrinsic conflict when patient empowerment encounters professional autonomy [17], which needs to be further investigated.

It appears that the overall trend in the results of this study is that a negative attitude predominated among the professionals. However, the difference between experienced and inexperienced nurses, as well as physicians, was significant. As the introduction of electronic access to patients' health records and other similar eHealth services is on the eHealth agenda of many countries, we find it important not only to display the attitudes of the majority of professionals to the idea of patient access to EHRs, but also to show the findings from real experience. Although the results from the pilot county are not strong enough to translate into real qualitative changes in attitude, we found the changes to be statistically significant and therefore worth discussion.

Again, the difference between experienced and inexperienced nurses as well as physicians, was significant. Possibly the nurses in the pilot county have not experienced an increased workload and no harm to the patients, which was stated as a fear by physicians in the free text responses [13]. Moreover, the eHealth service may have a positive effect on the patient consultation. The opinions of physicians who used the eHealth service for themselves or their relatives were remarkable. They differed significantly from physicians not using the service. They found that the eHealth service more important for patients and that it improved quality in care to a greater extent. Probably their own usage has a strong positive influence on their attitudes.

In previous studies on giving patients web access to their own health data, patients regard the online access as a step towards increased quality of care and patient empowerment [18,21]. Other studies indicate that patients think privacy concerns are outweighed by benefits [22] and physicians do not experience significance increase in work load [23]. Checkland et al. argue, somewhat in opposition to Freidson, that direct access to information over the Internet may contribute to enhanced trust in the patient-doctor relationship as the patient becomes knowledgeable and enlightened [19].

Information and communication technology innovations are being increasingly used to supply citizens with various health services which will lead to new demands on health and social care organizations. There is a strong need for further studies to investigate how this new trend will influence patients and health personnel. Knowledge that experienced staff hold a less negative attitude is a factor that could support future deployment processes and may lead to an increase in acceptance [24,25]. Therefore, it is important to disseminate these results to the inexperienced peers that today are standing on the verge of the deployment process of such eHealth services in their organizations.

It is in line with the interest of action researchers to address the challenge of overcoming the clash between patient empowerment and professional autonomy. One means could be to emphasize the introduction method of an innovation in a socio-technical way [18]. Different methods of measurement around the introduction of a system can be integrated with each other, e.g., TAM, Theory of Planned behavour and Innovation Diffusion Theory [24,25]. Yi et al. [24] offer a set of measurement criteria required for a specific intervention, which could be the deployment of a public eHealth service. Based on the suggestions from Yi et al. [24], an adaptation has previously been applied to support the introduction of an eHealth system in a Swedish county council, with successful results [18]. Such an adaptation would be interesting to repeat, when innovations like this eHealth service are to be implemented in other regions.

Conclusion

The main outcome of this study is that many professionals are reluctant when faced with patients having access to their EHR and that professionals with personal experience of the innovation are more positive as to the benefit for patients than healthcare professionals without experience. The study shows that health personnel have concerns about patients' direct access to their EHR in terms of the benefit for patients if their relatives also have access. Physicians and nurses having their own experience with online EHRs demonstrate a positive attitude towards the benefit for patients to a greater extent. The success of the introduction of reforms from the patients' point of view is accompanied by the possibility for professionals to get used to the eHealth services. We therefore recommend that healthcare professionals should be more involved in the implementation of Public eHealth services such as online EHRs and that real experiences of the professionals should be better disseminated to their inexperienced peers.

These findings, based on real experiences, are of interest in the upcoming introduction of similar eHealth services in Europe [25], as well as in other countries with a public eHealth strategy agenda.

Acknowledgements

The action research project "DOME - Deployment of Online Medical Records and eHealth Services" is financed by VINNOVA (2012-02233) – Swedish Governmental Agency for Innovation Systems. The SUSTAINS implementation project is supported by the European Commission (No 297206). We would also like to give our gratitude to the Uppsala County Council for supporting the distribution of the web questionnaires to the physicians and the Swedish Association of Health Professionals for distributing the questionnaire to their members. Special thanks to all the respondents who participated.

References

[1] Patient Data Act (SFS 2008:355) http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Patientdatalag-2008355_sfs-2008-355/ Retrieved 2014-12-21

[2] Ministry of Health and Social Affairs. National eHealth – the Strategy for Accessible and Secure Information in Health and Social Care, 2010. http://www.nationellehalsa.se/Content/Cms/img/temp/National%20eHealth%202010.pdf. Retrieved 2014-12-21

[3] European Commission, eHealth Action Plan 2012-2020 – Innovative healthcare for the 21st century,

http://ec.europa.eu/health/ehealth/docs/com_2012_736_en.pdf Retrieved 2014-12-21

[4] Freidson, Eliot. The Reorganization of the Medical Profession", *Medical Care Research and Review*, 1985: 42:11-35

[5] Kurunmäki, L. A hybrid profession – the acquisition of management accounting expertise by medical professionals, *Accounting, Organizations and Society,* 2004: 29, pp. 327-347.

[6] Eklund, B. and I. Joustra-Enquist. Sustains--direct access for the patient to the medical record over the Internet. *Stud Health Technol Inform* 2004: 100: 182-189.

[7] Ålander, T. Eklund, B. and I. Joustra-Enquist. Projekt Sustains – Journaltillgång över Internet. AllmänMedicin 2004: 6: 11-13.

[8] Lyttkens, L. SUSTAINS Intermediate trial evaluation – Uppsala, Report D6.1 v1.1, 14 Jan 2014. Available from: http://www.sustainsproject.eu/. Retrieved 2014-12-21.

[9] Lyttkens L. D6.4 v1.0 SUSTAINS Uppsala patient questionnaire. Uppsala County Council. Public Report. 2013-03-18

[10] Lyttkens L, Eklund B, Pettersson M. Rapport om införande av e-hälsotjänster. Min journal på nätet. Sustains. Landstinget i Uppsala län. Gov report 2014.

[11] Eklund B. Sustains, project leader. IT-dep.Uppsala County Council. Oral information 2015-04-02.

[12] Inera. Monthly report citizen eServices (in Swedish) Journalen. http://www.inera.se/TJANSTER--PROJEKT/Journalen/Inforande/ Last updated 2015-03-23. Retrieved 2015-04-03.

[13] Erlingsdottir, G., Lindholm C, Ålander T. (2014). eHealth services, patient empowerment and professional accountability – An empirical study on the changing patient-doctor relationship in the digital world. Conference paper, Edinburgh.

[14] Melin U., Axelsson K., and Lundsten M., Talking to, not about, Entrepreneurs – Experiences of Public e-service Development in a Business Start Up Case*, in eChallanges*, P. Cunningham, Editor. 2008

[15] Scandurra I, Åhfeldt, R-M, Cajander Å. Towards National Deployment of Medical Records and eHealth Services. *In Proceedings of VITALIS – Nordic leading eHealth meeting.* 2014: 14-17. Göteborg. Göteborgs Universitet; 2014. Retrieved 2014-12-21. https://gupea.ub.gu.se/bitstream/2077/35435/2/gupea_2077_35435_2.pdf

[16] Stata Statistical Software. Release 13.1 ed. College Station, TX: Stata Corporation 2014

[17] Erlingsdottir G, Lindholm C. When patient empowerment encounters professional autonomy. The conflict and negotiation process of inscribing an eHealth service. *Scandinavian journal of public administration*. 06/2015; 19(2)

[18] Hägglund M, I. Scandurra I, and Koch S, Empowering patients and relatives in elderly care by information access, in *the 6th International Conference on Information Communication Technologies in Health*, 2008, pp. 218–223.

[19] Checkland, K., Marshall, M. & S Harrison, 2004, Rethinking accountability: trust versus confidence in medical practice, *Qual Saf Health Care*. 2004; vol 13, pp 130-135 doi:10.1136/qshc.2003.00972

[20] Koch S, Improving quality of life through eHealth-the patient perspective., *Stud. Health Technol. Inform.*, vol. 180, pp. 25–29, 2012.

[21] Delbanco T, Walker J D, Darer, Elmore G, and Feldman H J. Open Notes : Doctors and Patients Signing On, *Ann. Intern. Med.*, vol. 153, pp. 121–125, 2010.

[22] Vodicka E, Mejilla R, Leveille S G, Ralston J D., Darer J D, Delbanco T, Walker J, Elmore, J. G. Online access to doctors' notes: patient concerns about privacy. *J Med Internet Res* 2013: 15(9): e208.

[23] Delbanco T, Walker J, Bell S K, Darer J D, Farag N, Feldman H J, Mejilla R, Ngo L, Ralston J D, Ross S E, Trivedi N, Vodicka E, Leveille S G Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. *Ann Intern Med* 2012: 157(7): 461-470.

[24] Yi M.Y, Jackson, J.D., Park, J.S., and Probst, J.C. Understanding Information Technology Acceptance by Individual Professionals: Toward an Integrative View, Information & *Management* 2006 (43:3), pp. 350-363.

[25] Venkatesh V, Thong J Y L, Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly 2012*: Vol. 36 No. 1 pp. 157-178.

[26] SUSTAINS: Support User Access to Information and Services http://www.sustainsproject.eu/ . Retrieved 2015-03-01

Address for correspondence

Ture Ålander, Läkarpraktik, Bergsbrunnagatan 1, SE-75323 Uppsala, Sweden. E-mail ture@tapraktik.se

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-158

Characterizing Patient-Generated Clinical Data and Associated Implications for Electronic Health Records

Elliot G. Arsoniadis[a, b], Rabindra Tambyraja[c], Saif Khairat[d], Cyrus Jahansouz[b], Daren Scheppmann[e], Mary R. Kwaan[b], Gretchen Hultman[a], Genevieve B. Melton[a, b]

[a]*Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[b]*Department of Surgery, University of Minnesota, Minneapolis, MN, USA*
[c]*Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA*
[d]*Carolina Health Informatics Program, University of North Carolina, Chapel Hill, NC, USA*
[e]*Fairview Health Services, Minneapolis, MN, USA*

Abstract

Patient-facing technologies are increasingly utilized for direct patient data entry for potential incorporation into the electronic health record. We analyzed patient-entered data during implementation of a patient-facing data entry technology using an online patient portal and clinic-based tablet computers at a University-based tertiary medical center clinic, including entries for past medical history, past surgical history, and social history. Entries were assessed for granularity, clinical accuracy, and the addition of novel information into the record. We found that over half of patient-generated diagnoses were duplicates of lesser or equal granularity compared to previous provider-entered diagnoses. Approximately one fifth of patient-generated diagnoses were found to meet the criteria for new, meaningful additions to the medical record. Our findings demonstrate that while patient-generated data provides important additional information, it may also present challenges including generating inaccurate or less granular information.

Keywords:
Patient generated healthcare data; Patient facing technology; Electronic Health Record;

Introduction

As patient-centeredness is an increasingly important tenet in healthcare, technologies engaging patients for individual assessment of clinical status, satisfaction, and education are being implemented in larger number. Moreover, the importance of patient engagement for attracting and maintaining patients is increasingly recognized by health systems [1]. As part of this, the electronic health record (EHR) has undergone significant change, including the integration of patient-facing portals with the EHR. Increasingly, portals and other patient-facing tools allow patients to access their own medical data and in some cases grant the ability to modify or comment on the data in these platforms [2].

A number of studies have shown that patients find the use of a patient portal for direct data entry to be a positive experience. Wu et al. reported 98% of patients in their large cohort found the use of a clinic-based kiosk for patient-generated family health history information to be useful [3]. A similar study found over 90% satisfaction with a similar online patient-facing family health history tool in a population of Veterans Administration patients [4]. Patient-facing tools relying on patient-generated data for realtime feedback have also been used with success in the areas of lifestyle modification [5] and chronic disease management [6].

To date, there are few studies assessing the quality of patient-generated healthcare data. While there have been some reports of positive effects with the addition of clinically relevant data [7,8], these studies have been limited to certain diagnoses, such as cardiac disease, or to certain portions of the medical record, such as family history [4,8]. In this study, we sought to better understand patient entered data broadly for different aspects of health history through an evaluation of the quality of patient-generated data using a patient-facing health history tool for past medical history (PMH), past surgical history (PSH), and social history (SH).

Materials and Methods

The patient-facing tool evaluated in this study was available September 2014 to patients with appointments at a surgery clinic for the University of Minnesota Physicians. The content of the tool was an organization-wide health history questionnaire approved by the medical directors of the practice plan. The tool was available to patients enrolled in the EHR online patient portal prior to the clinic appointment. For those patients not enrolled in the online patient portal or with portal access but who failed to utilize the tool prior to presentation in clinic, tablet computers were made available on site for completion prior to meeting with the provider. Previous studies have shown that patients respond well to tablet computer-based health applications, despite bias that these technologies might be unusable in certain patient populations, such as the elderly [9].

The questionnaires were only available in English and covered PMH, PSH, and SH. Patients were presented with a list of 40 common medical diagnoses, 22 surgical procedures, and four social history domains. PMH and PSH sections included "Yes" or "No" options, as well as a "Comments" section for free-text of additional conditions and procedures. The SH portion of the tool included questions related to tobacco use, alcohol consumption, illicit substance use, and sexual history. For the first three, patients were asked about use, type, amount, duration, and date of cessation (if applicable). The sexual history domain included current status, partner gender, and any birth control measures.

Following completion of the tool, results were immediately available to the clinician on the enterprise Epic EHR system. The provider then was presented with the opportunity to "accept" or "reject" patient entered information. If accepted,

the entries were then associated with discrete diagnosis terms and diagnosis codes in the PMH and PSH sections, as well as the discrete data fields of the SH section.

Table 1 – Physician-rater Diagnosis Coding

Physician-rater grade	Example
Duplicate Diagnosis Diagnosis already in PMH/PSH. Granularity of diagnosis is equal or near equal to that in the PMH/PSH.	Patient enters "Hypertension". "Hypertension" in PMH.
Duplicate Low Granularity Diagnosis Diagnosis already in PMH/PSH. Granularity of diagnosis is less that that in the PMH/PSH.	Patient enters "Kidney Problems". "Diabetic nephropathy" in PMH - Patient enters "Fracture or Broken Bone". "L3 compression fracture" in PMH.
New Low Granularity Diagnosis Diagnosis not in PMH/PSH and insufficient granularity to provide clinical utility.	Patient enters "Respiratory Problem". No previous respiratory diagnosis listed. Patient enters "Fracture or broken bone" but fails to mention anatomic location.
New False Diagnosis Diagnosis not in PMH/PSH and with sufficient granularity to provide clinical utility but on EHR review appears unlikely to be true.	Patient lists "Blood Clotting Disorder". Clinician notes "no evidence of thrombophilia" in Hematology clinic progress note.
New Diagnosis with Utility Diagnosis not in PMH/PSH and with sufficient granularity to provide clinical utility and corroborating EHR information.	Patient enters diagnosis of "Anxiety". No mental health problem is listed in PMH. Review of patient's medication list shows benzodiazepine. Patient enters a history of cholecystectomy. None is listed in PSH. CT scan report reads "gallbladder surgically absent".

We reviewed the data of 50 patients who completed the tool within the first six weeks of implementation, out of a possible 146 eligible participants (response rate 34.2%). Inclusion criteria included any new or return patient to the clinic who previously had been seen within the institution. A "New" visit type was designated for a visit where the patient had not been seen in surgery clinic before and were new to the provider, but had been previously seen by another provider (outside of our surgery clinic) within the Fairview Healthsystem. Since our analysis included comparing patient-generated data against data already contained within the enterprise EHR, patients new both to the clinic and to the healthcare system were excluded from this study. All patient-generated diagnoses were reviewed, regardless of acceptance by the provider.

As summarized in Table 1, each diagnosis entered was then reviewed by an independent physician-rater. The initial step required the physician to determine whether the patient-generated diagnosis was previously listed in the PMH and PSH of the patient chart. The next step involved gauging the level of granularity provided by the patient-generated diagnosis. For those diagnoses previously listed in the EHR, the rater then determined if the patient-generated diagnosis was either equal in granularity or lesser in granularity compared to the initial diagnosis. The physician-rater also gauged the granularity of any diagnosis deemed "New" on review in the initial step. Patient-generated diagnoses lacking sufficient granularity to enrich the clinical picture of the patient were classified as "New Low Granularity Diagnosis". Those patient-generated diagnoses judged to be of sufficient granularity to provide enrichment of the clinical history underwent a third level of scrutiny. In this final step, the physician-rater reviewed the patient chart to look for evidence to suggest if the patient-generated diagnosis was likely to be accurate. Clinical narratives from physicians, both from within and outside of the health system, were utilized as well as imaging studies, relevant laboratory values, medication lists, and procedure/endoscopy notes, as appropriate. Inter-rater reliability was assessed in five (ten percent) patient charts by a second physician-rater. Percent agreement and Kappa were calculated for coding of PMH/PSH diagnoses/procedures. Institutional review board approval was obtained and informed consent waived for this minimal risk study.

Results

Information about the patients and their types of visits are summarized in Table 2. In total, the 50 patients had a total of 435 patient-generated medical diagnoses, 231 surgical procedures, and 188 social history elements associated with these visits.

Table 2 – Patient Demographics and Visit Type

Demographic	n (%)
Age, range (mean)	18-74 (49.0)
Gender (female)	26 (52.0)
Visit Type	
New	6 (12.0)
Return	44 (88.0)
History Information	
Patient-generated diagnoses	435
Patient-generated procedures	231
Patient-generated social history	188

Over 50 percent of patient-generated medical diagnoses were duplicates of diagnoses already contained within the PMH portion of the EHR (Table 3). Of these duplicates, slightly greater than half of patient-generated diagnoses were of equal granularity to previous provider-entered diagnoses. The remaining half of the duplicated diagnoses had less granularity compared to previous provider-entered diagnoses. Nearly ten percent of the patient-generated diagnoses were determined to be new, but lacked sufficient granularity to provide meaningful clinical enrichment (New Low Granularity Diagnosis). Of the patient-entered diagnoses determined to be

both new and providing sufficient granularity to be considered clinically meaningful, nearly half were corroborated on chart review with the other half lacking sufficient evidence within the EHR to verify diagnosis (New Diagnosis with Utility and New False Diagnosis, respectively). Over 20 percent of new diagnoses deemed true based on chart review were related to mental health conditions (13.8% anxiety, 7.5% depression). History of colon/rectal polyps (11.3%) and history of blood transfusion (8.8%) were also common patient-generated diagnoses that were found to be true. Inter-rater reliability with a second physician-rater resulted in percent correlation of 0.89 and Kappa statistic of 0.817.

Forty-six of the 50 patients in this study completed the surgical history section of the survey. Two patients listed no previous surgeries. A total of 231 surgical procedures were entered by the remaining 44 patients (Table 4). Nearly 75% of surgical procedures were duplicates of procedures already recorded within the PSH portion of the EHR. Twenty-three percent were determined by an independent physician-rater to provide equal granularity to procedures previously listed. Nearly 52% were found to be of insufficient granularity compared to the surgical history already available in the EHR. The remaining surgical procedures entered were new to the PSH portion of the EHR. Of these, nearly half (ten percent of total) were determined to provide insufficient granularity to be of clinical utility. However, 11.3% of the total surgical procedures entered by patients were found to be new, sufficiently granular, and had evidence within the EHR. Three percent of surgical procedures entered by patients were new and sufficiently granular, but had no evidence within the EHR. One patient entered the same surgical procedure twice (Duplicate Entry – 0.9% of total). Inter-rater reliability with a second physician-rater resulted in percent correlation of 0.9 and Kappa statistic of 0.837.

Table 3 – Past Medical History

Category	n (%)
Duplicate Diagnosis	117 (26.9)
Duplicate Low Granularity Diagnosis	112 (25.7)
New Low Granularity Diagnosis	43 (9.9)
New False Diagnosis	82 (18.9)
New Diagnosis with Utility	80 (18.4)

All 50 patients completed at least one of the four domains within the SH portion of the survey. Overall, 188 data elements relating to tobacco, alcohol, and illicit substance use, as well as sexual history data, were entered. This patient-generated data for SH was compared to that already present within the EHR as depicted in Table 5. There were very few instances where patient-generated data did not concur with pre-existing data in the EHR. Only one new data element was found in both the tobacco and alcohol use portions of the SH domain. In the illicit substance portion, three instances of new information were noted. Interestingly, 11 instances of new information were observed for sexual history.

Table 4 – Past Surgical History

Category	n (%)
Duplicate Diagnosis	53 (23.0)
Duplicate Low Granularity Diagnosis	120 (51.9)
New Low Granularity Diagnosis	23 (10.0)
New False Diagnosis	7 (3.0)
New Diagnosis with Utility	26 (11.3)
Duplicate Entry	2 (0.9)

Results of physician ratings were compared between Past Medical and Surgical History diagnoses (Tables 3 & 4) using post-hoc chi-square analysis. Notable differences between medical and surgical diagnoses were observed in New False Diagnoses (18.9 vs. 3.0%) and New Diagnoses with Utility (18.4 vs. 11.3%) with significant results (χ^2 = 63.36, $p<0.0001$). Results were also compared by gender, age (≥50), and visit type in post-hoc analysis. When combining both medical and surgical diagnoses, we observed a difference in physician ratings between male and female patient-generated diagnoses, with the largest difference observed in the rate of New Low Granularity Diagnoses (13.7 vs. 6.6%, χ^2 = 11.17, $p=0.0247$). Similarly, the difference in physician ratings of diagnoses generated at New versus Return visit types approached significance (χ^2 = 9.35, $p=0.0529$), especially with Duplicate Diagnoses and New False Diagnoses (13.6 vs. 26.8% and 23.7 vs 12.4%). Age did not correlate with any difference in physician rating of patient-generated medical or surgical diagnoses (χ^2 = 2.42, $p=0.659$).

Table 5 – New Social History Data Elements

Category	n (%)
Tobacco	1 (5.9)
Alcohol	1 (5.9)
Illicit Substance	3 (17.6)
Sexual History	11 (64.7)
Total	17

Discussion

This study summarizes the quality and potential value of data generated with web and tablet-based data entry of health history information by patients. Previous studies have found clinical utility in patient-generated data related to chronic disease management, such as entering home blood pressure measurements using an online patient portal [6] and family history [8]. This study provides an analysis of more comprehensive patient-generated data.

Over half of patient-generated medical diagnoses and nearly 75 percent of surgical diagnoses were duplicates of those already recorded within the PMH and PSH portions of the EHR. This is consistent with other studies including Manaktala et al. and Okura et al which found good concordance between patient self-reported data and the medical record [7,10]. A broader study by Tisnado et al. examining a paper-based patient survey compared to the medical record found concordance between both data sources varied considerably. Concordance between diagnoses like "History of acute myocardial infarction" and "Diabetes" were high. Less concordance was evident with more subtle diagnoses such as "Depressed Mood", "High Cholesterol" or history of having had a previous echocardiogram [11]. Our results were similar to these previous studies. Similarly, much of the duplication of previously-recorded diagnoses in this study occurred with cardiovascular health (data not included in results).

Surgical diagnoses mirrored medical diagnoses, in that the vast majority of patient-generated data was a duplication of what already existed in the EHR. Interestingly, the rate of New False Diagnosis was considerably lower for PSH vs PMH (3.0 vs 18.9%). Likely, this is due to clearer knowledge on the part of patients regarding previous procedures. Patient self-knowledge of past surgeries is not a novel finding. In the previously mentioned work by Tisnado, the authors found excellent correlation between the patient survey and medical record for the diagnosis "History of coronary artery bypass or angioplasty" and fair agreement for previous cardiac catheterization [11].

The rate of correlation between patient and medical record is encouraging. More interesting, nearly 20 percent of patient-generated medical diagnoses and 11 percent of surgical ones were found to provide a new, sufficiently granular piece of data that was deemed to be likely accurate based on chart review by a content expert. These diagnoses represent discrete data elements not previously being captured in the PMH or PSH sections of the EHR. However, there was some evidence found within the clinical narratives, laboratory values, or imaging data contained within the institutional EHR. The most frequent patient-generated diagnoses that met criteria related to mental-health, specifically anxiety and depression. Hulse et al. found a similar phenomenon during the introduction of a patient-facing health history tool at their institution [12]. Previous literature comments on the prevalence of mental health diagnoses in patients with gastrointestinal disease [13]. The implications of uncaptured data regarding mental health conditions in this population of patients in a single clinic requires further inspection and analysis.

The failure to capture diagnoses as discrete data elements within the EHR has wide-ranging implications. As the EHR is increasingly utilized for secondary use, the need for complete data is more important. Tate et al. describes data quality as a 'sine qua non' for EHR use in randomized control trials [14]. Accurate and complete data is necessary to ensure high quality phenotyping, cohort discovery and recruitment into randomized controlled trials.

An additional 20 percent of patient-generated medical diagnoses and only three percent of surgical ones were deemed both new and sufficiently granular, but no evidence of disease or prior procedure could be found based on chart review. We were limited by the information provided by our institutional EHR. In this era of fragmented care, patients can have missing clinical data that is contained within other institutions or health systems. In some cases, our institutional EHR included clinical narratives, laboratory values, and the results of imaging studies performed outside. It is reasonable to assume that some portion of these patient-generated diagnoses are true and that supporting healthcare data exists within another institution's EHR .

Nearly ten percent of both medical and surgical diagnoses entered were new additions to the chart, but were of such low granularity that they were not clinically meaningful. Further, nearly half of the duplicated medical diagnoses and two-thirds of the duplicated surgical procedures were rated as less-granular versions of diagnoses or procedures already within the PMH and PSH sections of the EHR.

This study also gives an initial insight into the design of the patient-facing tool. Many of the diagnoses listed in the survey provided sufficient granularity without the need for free text entry by the patient (i.e. "Hypertension", "Diabetes mellitus", "Cholecystectomy"). In order to provide a comprehensive list of medical history conditions and surgical procedures that would not be found cumbersome for most patients, the survey was reliant on free text entry in order to provide sufficient granularity for certain diagnoses and procedures. For instance, when "Fracture or Broken Bone" or "Cancer" was selected, the tool relied on patient free text to specify the anatomic site or type. When "GI surgery" was selected, the tool relied on free text entry to specify the site and procedure. An overwhelming majority of patient-generated diagnoses did not include any free text entries associated with them. This was a major limitation of our study. It can be reasonably inferred that patients will not provide further granularity to diagnoses beyond that which is suggested to them on the Health History tool. Further work needs to be done to elucidate reasons why most patients are not engaging with the tool with free text entry. It may be that during the clinic intake process, patient-generated diagnoses from the tool need immediate addressing and clarification by clinic staff. This would require a major shift in intake workflow, most of which is currently being done by medical assistants. This also is an opportunity to reflect on the design of the tool, specifically in regards to the granularity of the diagnoses provided. Perhaps a drop-down menu approach would provide greater ease of use and ensure greater granularity compared to free text entry by patients. Another limitation was the low response rate of patients eligible to use the tool (34.2%). Clearly, buy-in from clinic personnel needs to be sought prior to introduction of these patient-facing technologies in order to address inevitable changes in workflow. This is being done in further iterations of this work.

One important limitation of the current study and an important area to further characterize is the amount of time required for providers to review patient-entered data. While there is an assumption that having patients enter their own data brings efficiencies, the time saved with patient-entry must be counter-balanced by the time to review and potentially correct and verify patient data. This question could likely be answered with an additional observational study or possibly a log file analysis of provider use of the EHR and consumption of patient entered data.

A significant portion of the patient-generated Social History data matched that which was already present within the EHR. There were very few instances within tobacco, alcohol, and illicit substance use history where patient-generated data was found to provide new information. Likely, this is related to the fact that well-established structured data fields have been in existence for these topics and have been utilized preferentially for some time over unstructured free text [15]. In the domain of sexual history, eleven instances of new information occurred. In nine of these instances, no previous sexual history had been recorded and the patient-generated data represented the first present within the EHR. Thirty nine charts (78 percent) contained previous provider-entered sexual history data. This mirrors previous work by Goyal et al. who observed 82 percent of patient charts containing a sexual history in their cohort of emergency department patients. We also found similar results in that when sexual history had been previously entered by the provider, it had a high rate of concordance with the patient-generated sexual history data [16].

Conclusion

Patient-generated health data is becoming more prevalent in healthcare. We are starting to see the EHR as a dynamic instrument in which patients not only access health data, but are engaged in data input as well. In this pilot study, we observed and graded patient-generated health data obtained during the rollout of a comprehensive health history tool via two patient-facing modalities. While the majority of patient-entered diagnoses were already captured in the EHR, nearly 20 percent of patient-generated medical diagnoses and eleven percent of surgical ones were found to be new, granular, and supported by other data within the chart. These results should be met with guarded optimism. Successful implementation of patient-facing technologies requires significant front-end work. IT experts, informaticists, and clinicians need to be engaged to design a suitable tool that combines ease-of-use with granular data capture. Provider and non-provider clinic personnel buy-in is also necessary early in the process, as these technologies will highly affect existing clinic workflows. Future work will include engaging both patients and clinicians to further optimize the user interface and workflows. Ultimately, it appears that patient-entered health history data has the potential of improving clinical data in the EHR and improving these approaches could potentially enable easy, accurate, granular data entry valuable to clinicians and researchers.

Acknowledgments

This research was supported by the University of Minnesota Department of Surgery, University of Minnesota Physicians, and Fairview Health Services.

References

[1] Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014 Jul-Aug;21(4):737-741.

[2] Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul-Aug;21(4):751-757.

[3] Wu RR, Orlando LA, Himmel TL, Buchanan AH, Powell KP, Hauser ER, et al. Patient and primary care provider experience using a family health history collection, risk stratification, and clinical decision support tool: a type 2 hybrid controlled implementation-effectiveness trial. BMC Fam Pract 2013 Aug 6;14:111-2296-14-111.

[4] Arar N, Seo J, Abboud HE, Parchman M, Noel P. Veterans' experience in using the online Surgeon General's family health history tool. Per Med 2011 Sep 1;8(5):523-532.

[5] Forjuoh SN, Ory MG, Wang S, des Bordes JK, Hong Y. Using the iPod Touch for Patient Health Behavior Assessment and Health Promotion in Primary Care. JMIR Mhealth Uhealth 2014 Mar 21;2(1):e14.

[6] Sun J, McNaughton CD, Zhang P, Perer A, Gkoulalas-Divanis A, Denny JC, et al. Predicting changes in hypertension control using electronic health records from a chronic disease management program. J Am Med Inform Assoc 2014 Mar-Apr;21(2):337-344.

[7] Manaktala S, Rockwood T, Adam TJ. Validation of pre-operative patient self-assessment of cardiac risk for non-cardiac surgery: foundations for decision support. AMIA Annu Symp Proc 2013 Nov 16;2013:931-938.

[8] Wu RR, Himmel TL, Buchanan AH, Powell KP, Hauser ER, Ginsburg GS, et al. Quality of family history collection with use of a patient facing family history assessment tool. BMC Fam Pract 2014 Feb 13;15:31-2296-15-31.

[9] Vawdrey DK, Wilcox LG, Collins SA, Bakken S, Feiner S, Boyer A, et al. A tablet computer application for patients to participate in their hospital care. AMIA Annu Symp Proc 2011;2011:1428-1435.

[10] Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol 2004 Oct;57(10):1096-1103.

[11] Tisnado DM, Adams JL, Liu H, Damberg CL, Chen WP, Hu FA, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? Med Care 2006 Feb;44(2):132-140.

[12] Hulse NC, Ranade-Kharkar P, Post H, Wood GM, Williams MS, Haug PJ. Development and early usage patterns of a consumer-facing family health history tool. AMIA Annu Symp Proc 2011;2011:578-587.

[13] Shah E, Rezaie A, Riddle M, Pimentel M. Psychological disorders in gastrointestinal disease: epiphenomenon, cause or consequence? Ann Gastroenterol 2014;27(3):224-230.

[14] Tate AR, Beloff N, Al-Radwan B, Wickson J, Puri S, Williams T, et al. Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. J Am Med Inform Assoc 2014 Mar-Apr;21(2):292-298.

[15] Chen E, Garcia-Webb M. An analysis of free-text alcohol use documentation in the electronic health record: early findings and implications. Appl Clin Inform 2014 Apr 16;5(2):402-415.

[16] Goyal M, McCutcheon M, Hayes K, Mollen C. Sexual history documentation in adolescent emergency department patients. Pediatrics 2011 Jul;128(1):86-91.

Address for correspondence

Genevieve B. Melton, M.D., M.A. Core Faculty, Institute for Health Informatics, and Associate Professor, Department of Surgery, University of Minnesota, MMC 450, 420 Delaware Street SE, Minneapolis, MN, 55455, USA.

Email: gmelton@umn.edu.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-163

Recognizing Clinical Styles in a Dental Surgery Simulator

Phattanapon Rhienmora[a], Peter Haddawy[b], Siriwan Suebnukarn[c], Poonam Shrestha[d], Matthew N. Dailey[d]

[a] *School of Science and Technology, Bangkok University, Thailand*
[b] *Faculty of Information And Communication Technology, Mahidol University, Thailand*
[c] *Faculty of Dentistry, Thammasat University, Thailand*
[d] *School of Engineering and Technology, Asian Institute of Technology, Thailand*

Abstract

Recognizing clinical style is essential for generating intelligent guidance in virtual reality simulators for dental skill acquisition. The aim of this study was to determine the potential of Dynamic Time Warping (DTW) in matching novices' tooth cutting sequences with those of experts. Forty dental students and four expert dentists were enrolled to perform access opening to the root canals with a simulator. Four experts performed in manners that differed widely in the tooth preparation sequence. Forty students were randomly allocated into four groups and were trained following each expert. DTW was performed between each student's sequence and all the expert sequences to determine the best match. Overall, the accuracy of the matching was high (95%). The current results suggest that the DTW is a useful technique to find the best matching expert for a student so that feedback based on that expert's performance can be given to the novice in clinical skill training.

Keywords:

Automated procedure assessment; Dynamic time warping; Virtual reality; Haptics; Dentistry.

Introduction

The teaching of surgical skills, which has traditionally followed an apprentice style approach, is now facing a number of challenges that are driving medical schools to seek alternatives. One challenge is that increasing enrollment combined with time constraints on the time of expert surgeons means that students typically do not receive as much supervised training as would be desirable. Rather, students end up spending a significant portion of their training time carrying out unsupervised practice. Furthermore, medical schools have little choice but to set requirements for surgical curricula in terms of amount of training time spent rather than in terms of level of proficiency achieved [1]. Even when assessment is carried out, it is by nature subjective, lacking sufficient standardization.

In response to these challenges, researchers have been developing new types of simulators and medical schools have been exploring the use of simulation in surgical training curricula. These efforts have particularly benefited from advances in virtual reality (VR) technology. VR simulators now exist for training in a variety of surgical domains, [2–6] and VR simulators have been introduced into dental curricula as training devices for clinical skill acquisition in several tasks [7, 8]. VR simulators enable collecting detailed data on relevant metrics throughout a procedure and thus provide a basis for detailed and objective procedure assessment. Indeed, a great volume of recent research has focused on developing metrics and techniques for objective assessment of surgical skill level [9–12]. Such analyses can be used for summative assessment and more excitingly for formative assessment by incorporating mechanisms for tutorial feedback generation into the simulators. But as research has progressed in this direction, researchers have come up against the complicating factor of the great variability with which any particular operation may be carried out [13, 14]. Such natural variability, not only in minute details of movement but also in style, poses a challenge for assessment. In domains where procedures may be carried out in multiple acceptable ways, an effective teacher must be able to recognize the style that a student is naturally using and provide corrective guidance relative to that style. This should also be the case for any simulator that generates tutorial feedback. To achieve this, the first challenge that must be solved is to recognize the style being used by a student even if it is being carried out incorrectly.

In this paper we present an approach to identifying the expert procedure from a set of stored procedures carried out in different styles that best match a student procedure. We make use of a VR dental simulator that we have developed for teaching dental procedures [6]. The system includes haptic feedback that can simulate tooth surface exploration and cutting for tooth preparation. It can automatically record performance and kinematic data on how experts and novices perform each step of a task, e.g., tool path, angulations and force used, which are not available in conventional skill training environments. For the present work we make use of only information about tool path and time.

Identifying stylistically similar procedures based metrics such as tool path and timing is challenging. Even if the expert and student are following the same basic path, the student movements are typically less smooth, resulting in differences in path location and path length. Velocity of movement of students is also typically slower than that of experts. Dynamic time warping (DTW) [15, 16] is a technique introduced for the analysis of handwriting that is designed to cope with some of these challenges. DTW determines the similarity of two trajectories of motion in cases where the trajectories may differ in terms of length and timing. In analysis of surgical procedures, DTW has been applied to the problem of spatio-temporal registration of trajectories for recognition of surgical gestures [17]. The present study aims to determine the potential of DTW in matching novices' tooth cutting

sequences with those of experts for the task of performing access opening to the root canals.

Methods

Dental virtual reality simulator and on-line data collection

We developed VR simulator that operates on a standard PC with an Intel Core 2 Duo 2.8GHz CPU, 4 GB RAM, and an nVidia GeForce 9600GT video card with 512MB of RAM connected to a 15-inch monitor and two PHANToM Omni haptic devices in a dual configuration, representing a handpiece and a dental mouth mirror. (Figure 1). The PHANTOM Omni haptic devices (SensAble Inc., USA) allow six degrees of freedom positional sensing and generate three degrees of freedom force feedback with a maximum of 3.3 N. The simulation software was developed with C++, OpenGL, and OpenHaptics SDK (HDAPI). The tooth model was acquired using three-dimensional micro-CT (RmCT, Rigaku Co., Tokyo, Japan). Tomographic images were obtained using comprehensive dental imaging software (i-VIEW, Morita Co., Tokyo, Japan). Three-dimensional reconstruction was performed using 600 of these two-dimensional images processed by volume rendering. During simulation, a user could feel different force feedback when cutting through enamel, dentin, and pulp.

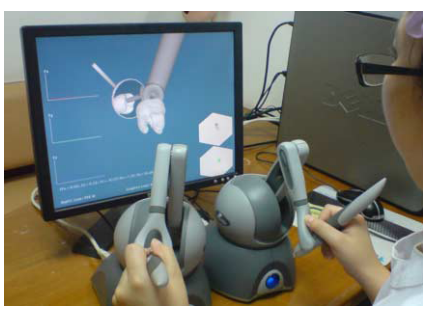


Figure 1 - A dentist performed access opening to the root canals with a haptic virtual reality simulator

Forty-four volunteers (40 fourth-year dental students (Novice), ages 20-23 years, and 4 dentists (Expert), ages 35-45 years) were enrolled in this study. All participants gave their written informed consent approved by the Institutional Ethical Review Board. Novices had experience using dental handpieces in cavity preparation from the operative preclinical course but no prior experience performing root canal treatment. Experts had professional training and experience in root canal treatment. All participants were right-handed. None of the participants had received any skill training using the haptic VR simulator.

All participants were briefly instructed on the use of the haptic VR simulator and the requirements of tooth preparation. The participants received a verbal explanation about the use of the system from the investigators and spent 15 minutes familiarizing themselves with the system interface, but not with the task. During this familiarization period, each participant was allowed to ask questions and receive further verbal explanation and suggestions from the investigators. After familiarization, the participants performed two trials of the task during which results were not recorded. The third trial was used for data collection. Participants were given the task of performing access opening to the root canals on the upper right first molar with the haptic VR simulator. The task was designed to mimic real access opening and to require hand-eye coordination for quality performance (Figure 1).

During data collection, data was gathered on elapsed time and kinematic variables with respect to the position of the handpiece in x, y and z direction, angulations of the handpiece with respect to x, y and z direction, transformation of the handpiece from the original position, drilling enabled/not enabled, position of the mirror in x, y and z direction, angulations of the mirror with respect to x, y and z direction, transformation of the mirror from the original position, and the force used in x, y and z direction.

Recognition of Clinical Process using DTW

Experts and novices performed operations on the dental simulator generating an output file which contained the kinematic attributes mentioned above in every millisecond. In this study, we selected the handpiece positions and movements as features for finding similarity between expert and novice and only the data when the user had enabled the drill for cutting was considered. Figure 2 shows the path of one expert and one novice in 3D space. Handpiece positions were normalized using $(X-\mu)/\sigma$, where X is the data sequence, μ is the mean and σ is standard deviation of those points.

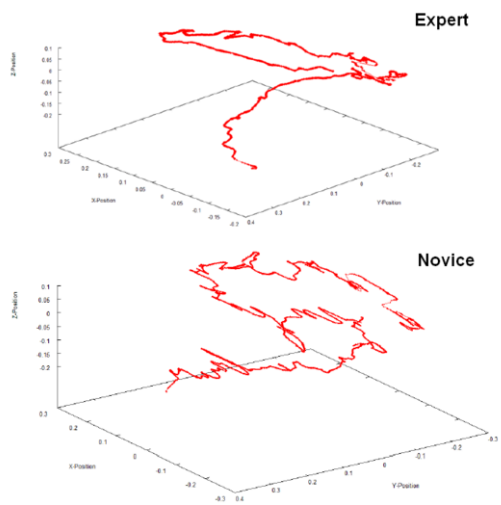


Figure 2 - Handpiece position in 3D generated from user's movement

Even when a novice is carrying out a procedure in the same basic way as an expert, the sequences of expert and novice movements will typically differ in length, positions, and speed. The expert's movements are usually more deliberate, economical and smooth throughout the operation. DTW is well suited to measuring similarity between sequences that may differ in speed and in which speed differences are not constant throughout the sequences. With DTW, the sequences are warped non-linearly in the time dimension to determine a measure of their similarity independent of non-linear variations in the time dimension but dependent on an application specific distance measure.

DTW [16] works as follows. Suppose we have two time series of length n and m respectively,

$$\text{Let } X = x_1, x_2, \ldots, x_n \text{ and } Y = y_1, y_2, \ldots, y_m$$

A matrix of size n by m is constructed such that the (i^{th},j^{th}) element of the matrix contains the distance $D(x_i,y_j)$ (in our case the Euclidean distance) between points i and j of the two series. After that, another matrix is constructed and the cost for each cell is determined using the condition:

- Initial Condition: $D(1,1) = 0$
- Recurrence: $D(i,j) = Dist(i,j) + min[D(i-1,j), D(i,j-1), D(i-1,j-1)]$

A warp path $W = w_1,w_2,...,w_k$ is constructed as the least cost path through the matrix from $(1,1)$ to (n,m) where k is the length of the warp path and the k^{th} element of the warp path is $w_k = (i,j)$. The warp path is typically subject to several constraints:

- Boundary conditions: The starting and ending points of the warp should be at two diagonally opposite ends of the matrix.
- Continuity: If $w_k = (a,b)$ then $w_{k-1} = (a',b')$ where $a - a' < 1$ and $b - b' < 1$. This restricts the allowable steps in the warp path to be adjacent cells (including diagonally adjacent cells).
- Monotonicity: If $w_k = (a,b)$ then $w_{k-1} = (a',b')$ where $a - a' > 0$ and $b - b' > 0$. This forces the points in W to be monotonically spaced in time.

Finally, cost is calculated as:

$$Cost = \frac{\text{Total cost of warp path}}{k}$$

The length of the warp path (k) in the denominator is used because for different sequences warp paths may have different lengths.

Data Analysis

Four experts performed access opening to the root canals using different sequences of operations.

Expert1 cut into the palatal pulp horn followed by extending the opening laterally to the mesiobuccal canal orifice, then followed a clockwise path to the distobuccal canal orifice, and the palatal canal orifice.

Expert2 cut into the palatal pulp horn followed by extending the opening laterally to the distobuccal canal orifice, then followed a counter clockwise path to the mesiobuccal canal orifice, and the palatal canal orifice.

Expert3 performed an initial cavity half way to the pulp chamber using a clockwise path then cut into the palatal pulp horn followed by extending the opening laterally to the mesiobuccal canal orifice, then followed a clockwise path to the distobuccal canal orifice, and the palatal canal orifice.

Expert4 performed an initial cavity half way to the pulp chamber using a counter clockwise path then cut into the palatal pulp horn followed by extending the opening laterally to the distobuccal canal orifice, then followed a counter clockwise path to the mesiobuccal canal orifice, and the palatal canal orifice.

Forty students were randomly divided into 4 groups of 10 students each. Students in groups 1, 2, 3 and 4 were trained to perform access opening to the root canals following experts 1, 2, 3, and 4 respectively.

Data from novices and experts was gathered as previously described and DTW was applied to determine whether each novice could be matched to the expert who had trained him based on the movement patterns, thus recognizing the approach to the procedure being attempted by the novice. For each novice procedure, DTW was applied to calculate the similarity between that novice and each of the four expert procedures. The lowest cost was taken as the best match. Figure 3 shows the warp between one expert and one novice sequence. Notice that these two sequences were warped from the beginning to the end.

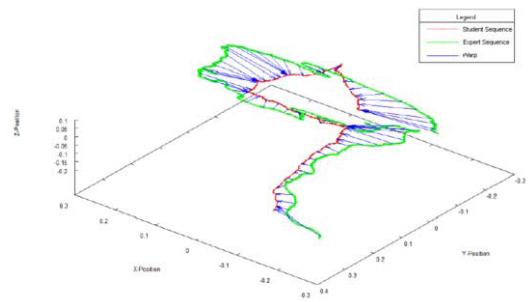


Figure 3 - Dynamic Time Warping between an expert and a novice sequence

The accuracy of the approach was computed as:

$$Accuracy = \frac{\text{Total number of correct matches}}{\text{Total number of students}} \times 100$$

Results

The data collection yielded 4 groups that differed widely for the tooth preparation sequence. Overall, the accuracy of the matching between novices and experts was very high (95%) compared to the probability of matching the sequences correctly by chance (25% in this four class problem). DTW correctly matched students in each group with the expert who trained them except for two cases. Case 1 was *Student8* (in group 1) who matched with *Expert2*, and case 2 was *Student24* (in group 3) who matched with *Expert1*.

In case 1, *Student8* should have matched *Expert1* but the lowest cost match was against *Expert2* at 170.45, with the match against *Expert1* being the next lowest at 339.54. The reason for this error can be seen from Figure 4. Figure 4(a) shows the sequences for *Student8* and *Expert1* and 4(b) shows the sequences for *Student8* and *Expert2*. The sequences in Figure 4(b) have initial segments that match closely, so DTW resulted in very low cost for the initial portion as compared to the sequences in Figure 4(a). Even though the cost was high for the later segment, the overall cost of the match between the sequences in 4(b) was lower than that in 4(a).

In case 2, *Student24* should have matched *Expert3* but the cost of the match against *Expert1* was lowest at 187.34 while the cost of the match against *Expert3* was only second lowest at 264.30. The reason for this mismatch can be seen from Figure 5. Figure 5(a) shows the sequences for *Student24* and *Expert3* while Figure 5(b) shows the sequences for *Student24* and *Expert1*. From Figures 5(a) and 5(b) it can be seen that the expert sequences were somewhat similar. The two expert sequences differ somewhat at the start but are similar at the

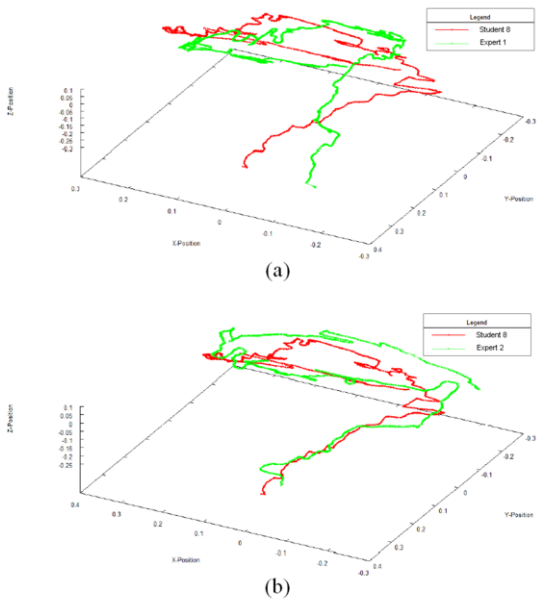


Figure 4 - Sequences of Student8 and Expert1 that should match (a) and sequences of Student8 and Expert2 that should not match but with lowest cost (b)

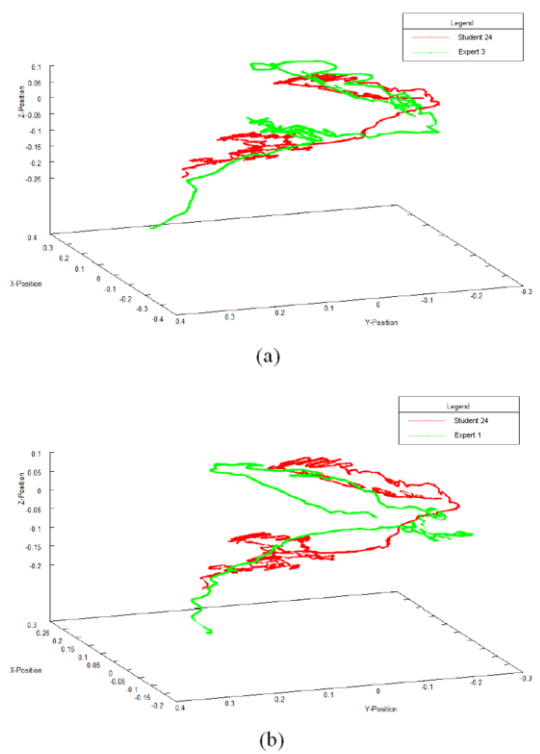


Figure 5 - Sequences of Student24 and Expert3 that should match (a) and sequences of Student24 and Expert1 that should not match but with lower cost (b)

later portions. Because of this, the costs of matching *Student24* with *Expert3* or *Expert* 1 were similar.

Discussion

Generally, knowledge of results (e.g., the pulpal depth was 1mm too deep) and knowledge of performance (e.g., was the handpiece held perpendicular to the long axis of the tooth when it should not have been) facilitate learning because they enable the performer to make appropriate adjustments to the movement on the next trial and increase the likelihood of the desired outcome will be achieved. In cases where procedures can be carried out using various alternative sequences, it is important to be able to recognize the correct sequence that a student is attempting in order to be able to provide effective corrective feedback. In this paper we have shown that metrics can be gathered from a virtual reality simulator that combined with dynamic time warping enable a student to be matched to the style of the expert who tutored him with a high degree of accuracy.

DTW seems to be a promising tool in the recognition of clinical performance in that it provides for objective evaluation of the trajectory of the instrument used to perform a dental procedure and is robust to differences in tool path length and velocity which differ greatly between novices and experts. The technique may enhance our understanding of the processes involved in clinical performance based on an unbiased assessment. Of course, the DTW technique alone cannot answer the question of how well a novice performs a task. Outcome measures need to be combined for comparison of outcomes under conditions where the movement trajectories are taxed, thus enabling analysis of their separate contributions to the outcome distortions in poor performance.

Conclusion

The current results suggest that DTW is useful technique to find the best matching expert for a student so that feedback based on that expert's performance can be given to the novice in clinical skill training. Several directions remain open for further work. First is the issue of improving the performance of the matching algorithm. In this paper we used only data of the tool location at each point in time. Performance could likely be improved by using information about trajectory of motion as well as information about other metrics such as force and angulation. The next step is to analyze the differences between the expert and novice procedures and identify those differences most relevant to quality of outcome for use in formative assessment. The final issue concerns how best to present this assessment information to the student to improve performance.

References

[1] Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, and Satava RM. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann. Surgery 2005: 241: 346-72.

[2] Riva G. Applications of Virtual Environments in Medicine. Methods Inf Med 2003: 42 (5): 524-34.

[3] Färber M, Hummel F, Gerloff C, and Handels H. Virtual reality simulator for the training of lumbar punctures. Methods Inf Med 2009: 48: 493-501.

[4] dV-Trainer, www.mimicsimulation.com/products/dv-trainer/, 2014.

[5] Delorme S, Laroche D, DiRaddo R, and Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training, Neurosurgery 2011: 71(1): 32-42.

[6] Rhienmora P, Haddawy P, Khanal P, Suebnukarn S, and Dailey MN. A Virtual Reality Simulator for Teaching and Evaluating Dental Procedures. Methods Inf Med 2010: 49(4): 396-405.

[7] Sohmura TH, Hojo H, Nakajima M, Wakabayashi K, Nagao M, Iida S, Kitagawa T, Kogo M, Kojima T, Matsumura K, Nakamura T, and Takahashi J. Prototype of simulation of orthognathic surgery using a virtual reality haptic device. Int J Oral Maxillofac Surg 2004: 33(8): 740-50.

[8] Steinberg AD, Bashook PG, Drummond J, Ashrafi S, and Zefran M. Assessment of faculty perception of content validity of PerioSim, a haptic-3D virtual reality dental training simulator. J Dent Educ 2007: 71(12): 1574-82.

[9] Megali G, Sinigaglia S, Tonet O, and Dario P. Modelling and evaluation of surgical performance using hidden Markov models. IEEE Trans Biomed Eng 2006: 53(10): 1911-9.

[10]Oropesa I, Sánchez-González P, Lamata P, Chmarra MK, Pagador JB, Sánchez-Margallo J, Sánchez-Margallo FM, and Gómez J. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res 2011: 171(1): 81-95.

[11]Suebnukarn S, Phatthanasathiankul N, Sombatweroje S, Rhienmora P, and Haddawy P. Process and Outcome Measures of Expert/Novice Performance on a Haptic Virtual Reality System. J Dent 2009: 37(9): 658-65.

[12]Overby DW and Watson RA. Hand motion patterns of Fundamental of Laparoscopic Surgery certified and noncertified surgeons. Am J Surg 2014: 207: 226-30.

[13]Reiley CE, Lin HC, Varadarajan B, Vagvolgyi B, Khudanpur S, Yuh DD, and Hager GD. Automatic recognition of surgical motions using statistical modeling for capturing variability, Studies in Health Technology and Informatics 2008: 132: 396-401.

[14]Anderson DK. How can educators use simulation applications to teach and assess surgical judgement? Academic Medicine 2012: 87: 934-41.

[15]Brina CD, Niels R, Overvelde A, Levi G, and Hulstijn W. Dynamic time warping: A new method in the study of poor handwriting. Hum Move Sci 2008: 27(2): 242-55.

[16]Kruskal J and Liberman M. The symmetric time-warping problem: From continuous to discrete. In: Sankoff D and Kruskal J, eds. Time warps, string edits, and macromolecules: The theory and practice of sequence comparisons. Massachusetts (USA): Addison-Wesley, 1983.

[17]Padoy N and Hager DG. Spatio-Temporal registration of multiple trajectories. In: Fichtinger G, Martel A, and Peters T, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, Lecture Notes in Computer Science 2011: 6891: 145-52.

Address for correspondence

Phattanapon Rhienmora, PhD

phattanapon.r@bu.ac.th

School of Science and Technology, Bangkok University

Rama 4 Road, Khlong Toei, Bangkok 10110, Thailand

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-168

Web-based auditory self-training system for adult and elderly users of hearing aids

Simone Virginia Vitti[a], Wanderléia Quinhoneiro Blasca[b], Daniel Sigulem[a], Ivan Torres Pisa[a]

[a] *Department of Health Informatics, Federal University of São Paulo (UNIFESP), São Paulo-SP, Brazil.*
[b] *Department of Audiology, University of São Paulo (USP), Bauru-SP, Brazil*

Abstract

INTRODUCTION. Adults and elderly users of hearing aids suffer psychosocial reactions as a result of hearing loss. Auditory rehabilitation is typically carried out with support from a speech therapist, usually in a clinical center. For these cases, there is a lack of computer-based self-training tools for minimizing the psychosocial impact of hearing deficiency. OBJECTIVE. To develop and evaluate a web-based auditory self-training system for adult and elderly users of hearing aids. METHODS. Two modules were developed for the web system: an information module based on guidelines for using hearing aids; and an auditory training module presenting a sequence of training exercises for auditory abilities along the lines of the auditory skill steps within auditory processing. We built aweb system using PHP programming language and a MySQL database .from requirements surveyed through focus groups that were conducted by healthcare information technology experts. The web system was evaluated by speech therapists and hearing aid users. An initial sample of 150 patients at DSA/HRAC/USP was defined to apply the system with the inclusion criteria that: the individuals should be over the age of 25 years, presently have hearing impairment, be a hearing aid user, have a computer and have internet experience. They were divided into two groups: a control group (G1) and an experimental group (G2). These patients were evaluated clinically using the HHIE for adults and HHIA for elderly people, before and after system implementation. A third web group was formed with users who were invited through social networks for their opinions on using the system. A questionnaire evaluating hearing complaints was given to all three groups. The study hypothesis considered that G2 would present greater auditory perception, higher satisfaction and fewer complaints than G1 after the auditory training. It was expected that G3 would have fewer complaints regarding use and acceptance of the system. RESULTS. The web system, which was named SisTHA portal, was finalized, rated by experts and hearing aid users and approved for use. The system comprised auditory skills training along five lines: discrimination; recognition; comprehension and temporal sequencing; auditory closure; and cognitive-linguistic and communication strategies. Users needed to undergo auditory training over a minimum period of 1 month: 5 times a week for 30 minutes a day. Comparisons were made between G1 and G2 and web system use by G3. CONCLUSION. The web system developed was approved for release to hearing aid users. It is expected that the self-training will help improve effective use of hearing aids, thereby decreasing their rejection.

Keywords:

Software (MeSH L01.224.900); Adult (MeSH M01.060.116); Elderly (MeSH M01.060.116.100); Hearing Aids (MeSH E07.814.458); Training (MeSH SP4.006.047.453.584); Rehabilitation (MeSH E02.831.800)

Introduction

Sensorineural hearing loss is a chronic disease that can have an impact on individuals' daily lives[1]. To minimize the day-to-day impact of such a disease, there is a need for hearing aids to be adapted for adults and elderly people with hearing loss. This population often does not adapt well to using hearing aids and these individuals often discard them. It has been shown[2,3,4] that auditory training helps users improve their discrimination and comprehension of speech sounds. Simply recalibrating the hearing aid does not help regarding users' auditory quality. Specific auditory training techniques are needed in order to positively influence the auditory processes of neuronal plasticity in these individuals.

In Brazil, there is no web-based system aimed towards adult and elderly users of hearing aids similar to the one proposed in this study which exists. This system provides an original application, and specific tools for auditory training, management and analysis. Likewise, there are no references in the literature relating to any system that addresses all of the skills relating to the mechanism and processes of the auditory system: sound detection, sound discrimination, auditory recognition, recognition of the temporal aspects of hearing and auditory performance in relation to competing acoustic signals and in unfavorable acoustic situations. This study is aimed to make it possible to ascertain whether an intervention through information technology (using a web-based system in the subject's home) enables the improvement of the user's auditory skills and adherence to hearing device use.

The objective of this study was to develop and evaluate a web-based system for the auditory training of adult and elderly users of hearing aids. This study formed part of the academic activities of the Health 360 research group (Saúde 360, saude360.com.br) a research group focused on consumer healthcare information technology, clinical decision support systems, e-health processes and evaluation, ontology for healthcare, knowledge discovery, data mining and big data. This group is part of the postgraduate program on Healthcare Management and Information Technology, Federal University of São Paulo (UNIFESP), Brazil.

Methods

The method used in this study was based on the construction and evaluation of a web-based system for the auditory training of adult and elderly users of hearing aids. This study was developed in the Department of Healthcare Information Technology, Paulista Medical School (EPM), UNIFESP, in partnership with the Department of Speech Therapy, University of São Paulo, Bauru (FOB-USP). Data was gathered in partnership with the Auditory Health Division (DSA) of the Craniofacial Abnormality Rehabilitation Hospital (HRAC-USP), in

Bauru. This study was approved by the Research Ethics Committee of EPM-UNIFESP, under procedural number 26949/2012, and the Research Ethics Committee of HRAC-USP, under the number 569743/2014.

The web-based auditory training system has two modules: (1) an information module based on guidelines for using hearing devices; and (2) an auditory training module based on different lines of auditory skills. Both modules were developed in PHP: Hypertext Preprocessor language (php.net). MySQL DBMS version 5.0 (mysql.com) was used for data storage and management.

The system was evaluated by five speech therapy specialists and two individuals with hearing deficiencies who were hearing aid users. They answered a questionnaire that evaluated layout requirements, content and system usability. These requirements were assessed based on a five level scale (very poor, poor, fair, good and excellent). The top three levels (fair, good and excellent) were considered to have indicated approval and that the two lowest levels rejection.

The population sample consisted of two groups of individuals who were selected according to the following inclusion criteria: age over 25 years; bilaterally symmetrical sensorineural hearing loss of mild to moderate degree; new users who had been bilaterally fitted with hearing aids featuring digital technology in accordance with the protocol of the Brazilian National Health System (SUS); literate; experience of web browsing; and available to undergo auditory training lasting 30 minutes a day for five days a week, over a four-week period. Initially, 100 individuals were selected. They were separated into a control group (G1) and an experimental group (G2). G1 did not have access to auditory training through the web-based system, while G2 had access to the auditory training module. A third experimental group named the web group (G3) was set up to assess system use and acceptance. G3 had access to both modules of the SisTHA portal, but the inclusion criterion established for this group was that the questionnaires should be answered before and after the auditory training. All the subjects in the groups needed to give responses to the Hearing Handicap Inventory for the Elderly (HHIE) questionnaire[5] or the Hearing Handicap Inventory for Adults (HHIA) questionnaire[6] and a questionnaire on hearing complaints, before and after application of the research protocol. The three groups are described and the functionality and usability of the web system are detailed below, in Figure 1. After the protocol had been applied, the results were analyzed. The structure of the method is shown below.

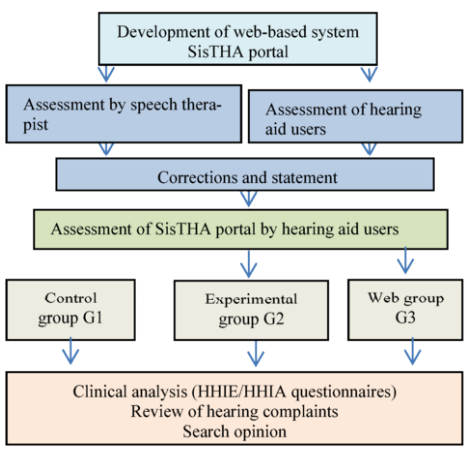


Figure 1 – Structure method for the study

Below is a diagram of the entity relationship model and database management system (DBMS) of MySQL Server ® version 5.0, furnishing a conceptual description of the data that was used to construct the SisTHA portal database (Figure 2):

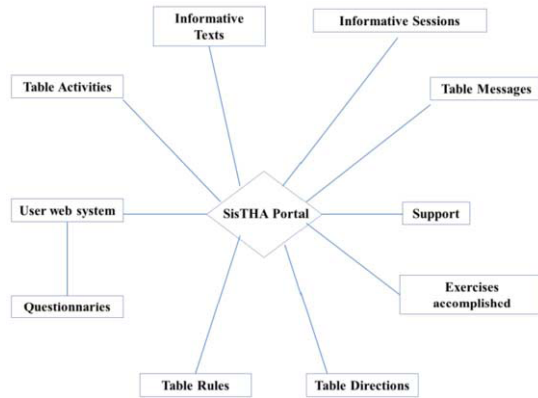


Figure 2 – Entity relationship model (ERM) of the SisTHA portal database.

Results and Discussion

The web-based system at www.sistha.com.br was finalized with two modules:

1) The information module had the aim of presenting general information about auditory health and was divided into two sections: hearing and hearing aids. These sections were composed of the following subsections: physiology of hearing; hearing loss; degree of hearing loss; the hearing aid; hearing test; hearing care; communication strategy; questionnaire on hearing complaints; importance of hearing aid; types of hearing aids; hearing aid and noise reduction; microphone of the hearing aid; battery; how to put the hearing aid on; mold and cleaning the hearing aid; how to use a telephone or mobile phone with a hearing aid. Figures 3, 4 and 5 shows the initial screens of the system relating to the information module.

Figure 3 - Screenshot showing the initial page of the information module.
(http://telemedicina6.unifesp.br/projeto/sistha).

Figure 4 - Screenshot of information module relating to the section on hearing aids and the subsection on types of hearing aid.
(http://telemedicina6.unifesp.br/projeto/sistha/index.php?cap=Aparelho%20Auditivo&tit=Tipos%20de%20aparelho%20auditivo)

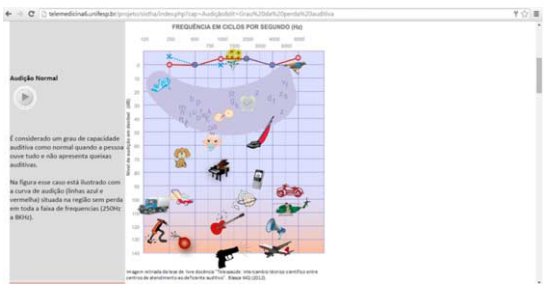


Figure 5 – Screenshot of information module showing the section on hearing aids and the subsection on degree of hearing loss.
(http://telemedicina6.unifesp.br/projeto/sistha/index.php?cap=Audi%C3%A7%C3%A3o&tit=Grau%20da%20perda%20auditiva)

2) The auditory training module was constructed by searching for and producing images and audio material that were coherent with each other and presented similarities to the day-to-day routines of the study population (in order to provide auditory training that would be as close as possible to the realities of the hearing aid user).

The auditory training was divided into five lines. Each of these corresponded to the stimulus of one of the mechanisms of the auditory skills of auditory processing: detection, discrimination, recognition, comprehension and auditory memory.

Each image of the auditory training module was correlated with corresponding audio material by means of the focus group method.[7,8] In this, 10 subjects presenting with normal hearing and vision assessed the images and audio material separately and then correlated them. Both the images and the audio material were selected only when there was a unanimous agreement within the focus group. The SisTHA portal presents 208 images in its database.

Audio recordings of voices were produced using the Audacity software. Recordings from twelve people were used (five men, five women and two children - a girl and a boy), with various accents and ages ranging from 9 years to 71 years. Each person was recorded while reading out numbers, words, sentences and text, which were all prepared in accordance with the daily lives of the adult and elderly subjects. The summary statistics for recorded material are listed below:

- 130 sentences;
- 200 individual words;
- 30 sequences of words, in two equal categories;
- 50 sequences of two words of different categories and lengths;
- 30 three-word sequences of equal categories;
- 50 sequences of three words of different categories and length;
- Numbers from 0 to 10;
- 40 three-digit number sequences;
- 20 texts composed of 200 to 600 words;
- Audio of news and vignettes.

The total amount of audio content produced for the content of the SisTHA portal was 10,027. However, all of this content was analyzed and selected. Through the analysis, 2,430 pieces of audio content were selected using audibility and intelligibility criteria. From these, some were selected to make up the audio overlay, in accordance with the proposal for auditory training through the SisTHA portal, for each line of the auditory training.

Currently, the audio database contains a total of 5,345 selected and finalized pieces of audio content.

The strategy of this module was to provide a sequence of exercises on the computer screen, to assist hearing aid users towards stimulating their hearing abilities. For this, the auditory training was divided into five lines. Each line corresponded to the stimulation of one of the mechanisms of hearing skills: detection, discrimination, recognition, comprehension and auditory memory.

Lines 1 to 4 of the auditory training presented three levels of difficulty (level 1 easy, level 2 medium and level 3 hard), while line 5, relating to communication strategy, had only one level of difficulty. For each level, a specific number of exercises were developed (Table 1), and were presented randomly to users, within their current stage (line and level).

Table 1– Number of exercises proposed within each line in the web-based system.

Stage	Total number of exercises	level 1	level 2	level 3
Line 1	120	40	40	40
Line 2	180	60	60	60
Line 3	156	16	70	70
Line 4	104	20	44	40
Line 5	15	15		
Line 6	Continuous/ random	Lines 1, 2, 3, 4, 5		

The difficulty of the auditory training increases as users correctly answers the exercises. When users give correct answers for 60% of the proposed exercises at each level, they move to the exercises at the subsequent level. After one line has been completed, users change stage, and so on. However, if they give wrong answers, the next exercise will remain at the level of difficulty of the previous exercise (which will be drawn within the same line and level). A running total of the number of correct answers is kept, with the aim of avoiding user frustration. It should be noted that the web-based system is dynamic and that exercises can be added at any time that is convenient.

Figures 6, 7, 8 and 9 present the screens of the SisTHA portal with examples of exercises within different lines and at different levels.

Figure 6 - Screenshot of the SisTHA portal referring to an example of an exercise within line 1 at level 1. By clicking on the "play" button, audio content is presented to the user, who has to choose the correct image associated with the audio.

Figure 7 - Screenshot of the SisTHA portal referring to an example of an exercise within line 1 at level 2. By clicking on the "play" button, audio content is presented to the user, who has to choose the correct image associated with the audio: the voice of a man, woman or child.

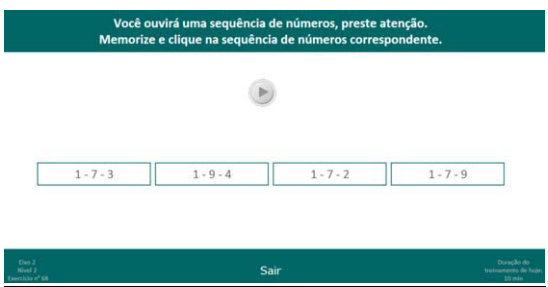


Figure 8 - Screenshot of the SisTHA portal referring to an example of an exercise within line 3 at level 3. By clicking on the "play" button, audio content is presented to the user, who has to choose the correct image associated with the audio, with regard to comprehension and auditory memory.

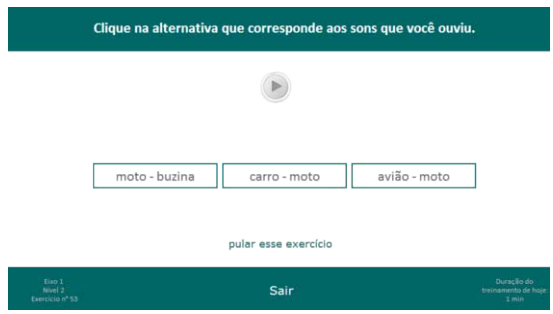


Figure 9 - Screenshot of the SisTHA portal referring to an example of an exercise within line 2 at level 1. By clicking on the "play" button, audio content is presented to the user, who has to choose the correct image associated with the audio.

As described earlier, the SisTHA portal was assessed by five volunteer specialist speech therapists with between 15 to 28 years of professional experience in the field of auditory rehabilitation. These professionals had academic qualifications as specialists in speech therapy, alongside their master's and doctoral degrees. They were working in universities and consultation offices. Evaluation of the system was also performed by two volunteer hearing aid users with bilateral hearing loss and a minimum of 25 years of hearing aid use. This evaluation consisted of a questionnaire which assessed the layout, content and usability requirements of the system. Responses were coded via a five-level scale (very poor, poor, fair, good and excellent). The top three levels (fair, good and excellent) indicated approval and that the two lowest levels indicated rejection. Responses to the questionnaires evaluating the system were analyzed by using descriptive statistics and content analysis from the perspective of Bardin.[9]

From the point of view of the speech therapists, the information module was approved with regard to the access requirements, layout, content, images, audio content and videos. The auditory training module was approved with regard to access, layout, volume adjustment, images, audio content and coherence between image and audio. Regarding the impact of the web-based system for hearing aid users, all the experts stated that the expected impact was positive. The evaluators reported that the following were important suggestions for the web system: improvement of the presentation of images, given the delays in loading them; insertion of a screen at changes in line regarding the number of correct answers per level; insertion of an incentive screen; and provision of access to screening of auditory processing before and after the auditory training to ascertain whether there was any improvement in the user's hearing abilities.

Regarding the evaluation of the web system by hearing aid users, both modules (information and auditory training) were approved in relation to all of their requirements. The hearing aid users provided a suggestion for the web system that an alternative should be included in the exercises such that users could choose not to answer the question (this option is already included in the current version). In addition, it was suggested that both a screen showing the number of correct answers per line and per level and a screen showing the change in line should be inserted.

Currently, the web-based system is at the data-gathering phase with the control, experimental and web groups until May 2015. Management and analysis of results will be carried out in June 2015.

Conclusion

A web-based system for helping in training the auditory skills of hearing aid users was constructed and evaluated for clinical use. In future clinical trials using the control, experimental and web groups, it is expected to be possible to assess whether standardized training done at home enables effective use of hearing aids, thereby giving rise to the possibility of reducing their rejection.

Acknowledgement

To FAPESP for assistance through the doctoral scholarship granted (number 2012 / 05410-0); to CNPQ for its Universal Notice with procedural number 4809 / 2011-8 in the name of Prof. Dr. Wanderléia Q. Blasca; and to the Auditory Health Division of HRAC-USP and the Department of Health Informatics, EPM-UNIFESP, for support and encouragement in the development of this study. We especially thank the speech therapists and hearing aid users who contributed to this study.

References

[1] Smith SL, Kricos PB. Rehabitation for the elderly patient. Advance for Audiologists. 2002; 4:35-37.

[2] Sweeton RW, Sabes JH. Tecnologic advances in aural rehabilitation: applications and innovative methods of service delivery. Trends Amplif. 2007; 11(2):101-11.

[3] Sweeton RW. Training the auditory brain to hear. The Hearing Journal. 2005; 58(6):10-16. processing disorder. Rev. soc.bras. fonoaudiol. 2007;12(4):310-14. Availabe at: http://www.scielo.br/pdf/rsbf/v12n4/v12n4a10.pdf

[4] Zalcman TE, Schochat E. The efficacy of formal auditory training in individuals with auditory

[5] Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. Ear Hear. 1982;3(3):128-34.

[6] Newman CW, Weinstein BE, Jacobson GP, Hug GA. The hearing handicap inventory for adults: psychometric adequacy and audiometric correlates. Ear Hear. 1990;11(6):430-3

[7] Dias CA (2000). Grupo Focal: técnica de coleta de dados em pesquisas qualitativas.10:2. Disponível em: http://periodicos.ufpb.br/ojs/index.php/ies/article/view/330/252.

[8] Drabenstott K M (1992). Focused group interviews In: Glazier J D, Powell RR. Qualitative research in information management. Englewood, CO: Libraries Unlimited. p. 85-104

[9] Bardin L. Content analysis. Lisboa: Edições 70, 1977.

Address for correspondence

Simone Virginia Vitti, Postg**raduate Program on Health Informatics, Department of Health Informatics, Federal University of São Paulo,** Rua Botucatu 862, Edifício Leal Prado, térreo, **São Paulo, Brazil.**

Telephone: +551155764347

Cell phone: +5512981165569

E-mail: simone.vitti@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-173

Design, Implementation and Evaluation of an Architecture based on the CDA R2 Document Repository to Provide Support to the Contingency Plan

Fernando Campos [a,c], Daniel Luna [a], Dean F. Sittig [b], Fernán González Bernaldo de Quirós[a]

[a] *Health Information Department, Hospital Italiano de Buenos Aires, Argentina*
[b] *School of Biomedical Informatics, University of Texas Health Sciences Center at Houston, Houston, TX, USA*
[c] *HL7 Argentina*

Abstract

The pervasive use of electronic records in healthcare increases the dependency on technology due to the lack of physical backup for the records. Downtime in the Electronic Health Record system is unavoidable, due to software, infrastructure and power failures as well as natural disasters, so there is a need to develop a contingency plan ensuring patient care continuity and minimizing risks for health care delivery. To mitigate these risks, two applications were developed allowing healthcare delivery providers to retrieve clinical information using the Clinical Document Architecture Release 2 (CDA R2) document repository as the information source. In this paper we describe the strategy, implementation and results; and provide an evaluation of effectiveness.

Keywords:

Electronic Health Record; CDA repository; Contingency plan.

Introduction

There is an increasing use of electronic health record, specifically the replacement of paper-based health record with Electronic Health Records (EHR) [1]. Therefore, every healthcare delivery process relies on information systems to ensure patient care. This brings up the threat of a lack of information for decision making in case of system downtime. One of the biggest risks in this scenario is inaccessibility of information needed to administer medication [2, 3].

Unexpected information technology (IT) downtime occurs more and more often with widespread adoption of electronic systems in healthcare [4]. Although the reliability of computing hardware has improved significantly, the complexity of software has escalated, especially in healthcare [4]. Risk identification and risk assessments are essential steps for developing preventive measures. Equally important is institutionalization of contingency plans as our data show that not all failures of health IT can be predicted and thus effectively prevented [5].

Most institutions had only partially implemented comprehensive contingency plans to maintain safe and effective healthcare during unexpected EHR downtimes [6]. Preparing for these unexpected downtimes should be a part of every EHR-enabled healthcare organization's overall patient safety strategy.

As most experts state, an effective contingency plan should address the causes and consequences of EHR unavailability, triggering processes and preparations that can minimize the frequency and impact of such events, ensuring continuity of care [7-9].

The objective of this paper is to describe the design, implementation, and evaluation of the IT component of a contingency plan that uses the Clinical Document Architecture (CDA) Release 2 (R2) document repository to support continuity of care during system downtime.

Methods

Settings

The Hospital Italiano of Buenos Aires (HIBA) is a nonprofit academic medical center founded in 1853. HIBA has a network of two hospitals with 750 beds (250 for intensive care), 800 home care patients under care, 25 outpatient care centers, and 41 operating rooms. There are more than 2,800 physicians, an equal number of medical team members and 1,900 administrative and support staff. During 2013-2014 there were 45,000 admissions, 3 million outpatient visits and 45,000 surgeries (half of them ambulatory).

In 1998 HIBA began the gradual implementation of a Healthcare Information System (HIS) developed in-house, from data capture to analysis. It includes a web based, modular, problem-oriented and patient-centered EHR. This EHR is known as Italica and allows inpatient, outpatient, home care, and emergency care records. Italica also allows users to order ancillary tests, prescribe medications and view results including imaging through an integrated picture archiving and communications system (PACS).

The EHR has a relational database record and also a CDA R2 document-based repository, which is digitally signed by professionals participating in healthcare delivery. This repository is used to interoperate with payers and other EHRs, and to make information portable for patients or other external healthcare providers. The system architecture for sharing medical information is based on HL7 CDA and a repository/registry XDS (Cross Enterprise Document Sharing) model defined by IHE (Integrating the Healthcare Enterprise). The document repository currently contains 36.4 million CDAs [10].

This repository allows the organization to operate without need for paper records, because information exchange between actors or systems is facilitated by these documents. For instance, for ancillary systems like imaging or laboratory, order is no longer paper-based but a digitally signed CDA R2. Likewise, result reports are not printed, but reviewed directly in the EHR through a CDA R2 sent by the ancillary service.

Since 2012, and based on this implementation, all procedures and communications for healthcare continuity while systems recover from a meaningful interruption have been redesigned, mainly for the inpatient setting. Two levels of contingency were identified:

- Level 1 is the application level, when only the EHR is not available. Causes might be a problem with the deployment of a new version or server issues. The rest of the computing infrastructure is available: database, networking, electricity, etc.
- Level 2 is total impact. Usually this level of contingency occurs when the database server is down or halted during upgrades or maintenance, or during data center problems affecting the server farm or storage, network outages or natural disasters.

The decision was to leverage the redundancy generated by the document repository to support contingency processes.

Design

A laboratory function study was performed to evaluate effectiveness and 20 Level 2 contingency tests were run on different days and at varying times. Time was randomized from the last execution to the start of simulated system downtime.

This involved simulating system downtime on 20 occasions at different times of day and comparing each prescription that each patient had at the time of the simulated downtime with the current list of prescriptions in real time. For example, if the list had been generated at 12:00 am and the system went down at 12:16 am, prescriptions during those 16 minutes were compared for additions or modifications.

Results

In order to mitigate the first level of contingency, a CDA navigator was developed, having as indexes some of the elements in the CDA header (metadata).

Table 1

Document date	/ClinicalDocument/effectiveTime/@low
Document type	/ClinicalDocument/code/@code
Service	/ClinicalDocument/legalAuthenticator/assignedEntity/representedOrganization/id/@extension and /ClinicalDocument/legalAuthenticator/assignedEntity/representedOrganization/id/@assigningAuthorityName
Patient id	/ClinicalDocument/recordTarget/patientRole/id/@extension
Patient root	/ClinicalDocument/recordTarget/patientRole/id/@root
First name	/ClinicalDocument /recordTarget/patientRole/patient /name/@given [1]
Last name	/ClinicalDocument /recordTarget/patientRole /patient/name/@family

Using this index, a tree is generated, and this tree is accessed based on patient information. From the tree root (target patient), timeline for inpatient (day by day) can be navigated. After selecting a specific date, the caregiver can access all documents for the visit grouped by document type and service.

This application is deployed on a different server from the EHR, and with a different and redundant database. If by any chance the EHR is not available, the document-based EHR at least can be retrieved.

Figure 1 – CDA Navigator

The second level of contingency assumes a total lack of database support, application server support, and electricity or network connections.

In this extreme case and only for inpatient and emergency care (average 720 beds), the question of which information is most critical for healthcare delivery to minimize risk to patients was evaluated on a consensus basis.

In this regard, two elements were key: access to the patient's prescriptions (medications, dosages, etc.) and proper labeling for laboratory samples.

Based on this requirement, an application was designed that accessed the document repository every 30 minutes for the latest medication and sampling label information for each inpatient.

Each computer running the application devotes local disk space to a folder tree organized by department and inpatient location and containing every prescription CDA and label information for each location.

This temporal repository is composed of two folders. The system alternates storage of the document between them. This application runs redundantly on several computers in strategically selected locations, and can only be used in contingency situations. These computers have a local printer and connection to backup power, with uninterruptible power source (UPS) and supplies of printer paper.

In the event of natural disasters, application server downtime, database downtime, or any other contingency, these computers can be used to print prescriptions and labels so that critical healthcare delivery can continue uninterrupted.

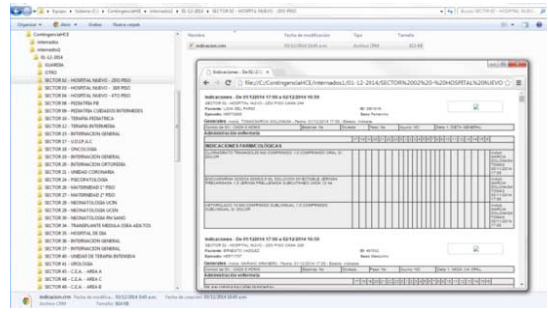


Figure 2 – File Explorer for Level 2 Contingency

Creating this backup requires querying the relational system and generates an average of 18 queries involving 44 tables for each patient. With an average of 720 beds, 12,960 queries would be needed every 30 minutes, which then would need to be turned into HTML for on screen rendering.

Using the CDA repository, only 1 query is needed, involving a snapshot of the current active inpatients or emergency care episodes, and the query to the document repository for the URL for the patient's last prescription CDA. Rendering the information only requires transformation of the XML using an XSLT stylesheet.

The downtime record between March 2012 and June 2014 is presented in Table 2. This table shows dates and times when the EHR was not available, and for how long. The column 'Type' indicates if the lack of EHR was programmed or not, and then the cause. The column "Level" indicates the severity level assigned to the cause and for each event, the quantity of pages printed.

Table 2 – Downtime Record and Pages Printed

Date Begin Time	Duration (Hrs.)	Type *	Motive	Level	Pages
08/03/12 09:15	05:55	U	EHR bug	1	365
05/04/12 23:10	02:00	U	DB issue	2	301
07/04/12 20:00	02:00	P	DB maintenance	2	10
27/04/12 09:50	06:15	U	EHR bug	1	303
12/05/12 22:40	07:50	P	Upgrade DB	2	1254
09/06/12 22:00	07:00	P	Upgrade Server memory	2	867
29/12/12 05:30	04:30	U	EHR bug	1	425
23/04/13 05:30	25:01	U	Shut-down power	2	1450
26/04/13 12:30	01:00	U	Router down.	2	23
27/04/13 18:00	03:00	P	Server maintenance	2	150
25/05/13 16:00	05:05	P	Server maintenance	2	1050
15/06/13 17:00	07:38	P	Server maintenance	2	1080
26/07/13 20:00	13:30	P	New DB cluster	2	1250
14/09/13 18:00	02:00	P	Update switches firmware	2	306
11/01/14 02:00	04:30	P	Update switches firmware	2	309
24/06/14 12:10	01:10	U	DB issue	2	124

*Type: U: Unplanned - P: Planned

As an example, a change of servers' operating system produced downtime of 13:30 hours. In this contingency, all prescriptions were printed, and 1,250 sheets of paper were used to print 759 prescriptions in 1 hour 15 minutes.

In the contingency Level 1 case, even though the CDA browser was used, the critical care units and emergency care areas preferred to print out the prescriptions in order to be able to make written notes of instructions and any other changes.

Details of each test simulation are presented in Table 3, including the time elapsed since last generation (mins), the number of inpatients at that moment, records that matched at the simulated downtime and the differences found.

Table 3 - Evaluation Cases

Case #	Minutes elapsed since last generation	Inpatients at down-time	Matches	Diffe-rences
1	2	654	650	4
2	10	695	667	34
3	18	712	668	44
4	25	730	666	64
5	5	759	723	36
6	12	759	707	52
7	26	762	686	76
8	5	771	759	12
9	5	721	697	24
10	15	678	630	48
11	14	699	643	56
12	10	712	680	32
13	17	734	686	48
14	5	723	707	16
15	15	745	709	36
16	13	711	671	40
17	22	733	657	76
18	5	732	712	20
19	1	722	722	0
20	25	745	673	72
		14,497	13,713	784

The differences were evaluated in each case: how many were caused by a patient being admitted or discharged, how many were caused by a change in medications. Medications for newly admitted patients and missing updates were deemed errors. New inpatients with no medications and discharged patients still in the contingency system were not deemed errors.

For example, for the second case, 34 differences were found, 22 of them due to patients admitted during the 10 minutes from the listing generation but with no prescription, 6 patients with their prescriptions updated, 13 patients discharged and 3 patients had new prescriptions. The evaluation result is presented in Table 4.

Table 4 – Evaluation Results for Contingency Listing

Case #	Diff.: New Admission	(1) Diff. Medication Change	Diff.: Discharged patient	(2) New patient with prescriptions	Errors (1) + (2)
1	1	1	2	0	1
2	12	10	6	6	16
3	22	6	13	3	9
4	34	10	17	3	13
5	18	3	15	0	3
6	26	4	20	2	6
7	8	10	55	3	13

8	6	3	3	0	3
9	12	3	9	0	3
10	22	2	20	4	6
11	29	6	19	2	8
12	16	5	11	0	5
13	23	10	6	9	19
14	8	2	6	0	2
15	17	5	12	2	7
16	20	7	13	0	7
17	39	17	16	4	21
18	10	3	7	0	3
19	0	0	0	0	0
20	36	9	23	4	13
	359	116	273	42	158

The results show that printed prescriptions concur in 98.91% (14,339/14,497) of those registered in the EHR.

The relationship between the time elapsed since the last generation and the differences found after downtime are presented in Figure 3.

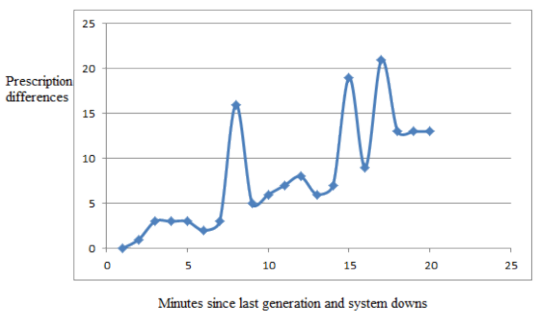


Figure 3 – Differences (Time)

Discussion

This architecture allows information to be available from all delivery care locations, and allows every actor requiring the information to access it.

The printing process has a predefined sequence for the various locations, prioritizing critical care units, emergency care, pediatric care, and then other services.

The more time elapses between the generation of the latest list and the EHR unavailability, the greater the difference in the prescription record. Although the medical staffs understand that any changes made in less than 30 minutes could be re-prescribed, they are aware that the nurse work based on the lists received and these should be corrected manually after the EHR disruption.

In the real scenario, other times should be evaluated because doctors do not make decisions on every patient immediately after the system goes down. Other changes could have been made during the downtime and will not be represented in the print version (prescriptions, bed changes, etc.). So we have two mechanisms that generate gaps between the information documented and the reality: gaps between the last local backup and the down time and gaps between the down time and the moment at which the caregiver uses the print version.

This study evaluated the differences between Point 1 and Point 2. The effectiveness of this architecture at Point 3 is pending evaluation and will be the subject of a future study.

The alternating folder strategy was required because in one of the tests if network connectivity or power was interrupted just when the folder was being generated, its structure could be corrupted. If this happens, the other folder can be accessed.

Prescription printed forms are used mainly by the nurses to ensure care continuity for patients. They continue recording their actions there in order to update the EHR when the system comes back, or scanned as part of the care record.

One limitation, as shown by the evaluation of the results, is that not all the information is available because there can be changes between the executions of the refresh process: patient transfers, new prescriptions, admissions or discharges.

No other experience in creating this kind of repository was found in the literature, either as a redundant repository or for use on a contingency basis.

Conclusion

Downtime in information systems, both planned and unplanned, is inevitable, even in the healthcare environment, where systems need to be available 24-7-365. Implementing tools such as the one presented in this paper provides contingency plan support and helps mitigate many risks that threaten information availability at the point of care.

References

[1] Wright A, Henkin S, Feblowitz J, McCoy AB, Bates DW, Sittig DF. Early results of the meaningful use program for electronic health records, N. Engl. J. Med. 2013;368:779–780. http://dx.doi.org/10.1056/NEJMc1213481.

[2] Hanuscak TL, Szeinbach SL, Seoane-Vazquez E, Reichert BJ, McCluskey CF. Evaluation of causes and frequency of medication errors during information technology downtime. Am J Health Syst Pharm. 2009;66(12):1119-1124.

[3] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors . JAMA. 2005;293:1197-1203.

[4] Lei J, Guan P, Gao K, Lu X, Chen Y, Li Y, Meng Q, Zhang J, Sittig DF, Zheng K. Characteristics of health IT outage and suggested risk management strategies: an analysis of historical incident reports in China. Int. J. Med. Inform. 2014;83(2):122-130.

[5] Sittig DF, Singh H. Electronic health records and national patient-safety goals. N. Engl. J. Med. 2012;367(19):1854–1860.

[6] Sittig DF, Gonzalez D, Singh H. Contingency planning for electronic health record-based care continuity: a survey of recommended practices. Int J Med Inform. 2014;83(11):797-804. doi: 10.1016/j.ijmedinf.2014.07.007. Epub 2014 Aug 7.

[7] Kilbridge P. Computer crash – lessons from a system failure. N. Engl. J. Med. 2003;348:881–882.

[8] Merrick J. 'Serious' computer crash hits hospital trusts. Daily Mail (July 31, 2006) 3984, Available at: http://www.dailymail.co.uk/news/article-77/Serious-crash-hits-hospital-trusts.html.

[9] Sittig DF, Ash JS, Singh H. The SAFER guides: empowering organizations to improve the safety and effectiveness of electronic health records. Am J Managed Care. 2014;20(5):418-423.

[10] Campos F, Plazzotta F, Luna D, Baum A, de Quirós FG. Developing and implementing an interoperable document-based electronic health record. Stud Health Technol Inform. 2013;192:1169. PubMed PMID: 23920943.

Address for correspondence

MSc. Fernando Campos
Chief of Software Engineering
Health Informatics Department
Hospital Italiano de Buenos Aires
HL7 Argentina Chairman
fernando.campos@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-178

Archetype based patient data modeling to support treatment of pituitary adenomas

Claire Chalopin[a], Dirk Lindner[b], Stefan Kropf[a], Kerstin Denecke[a]

[a] *Innovation Center Computer Assisted Surgery, Universität Leipzig, Germany*
[b] *University Hospital, Department of Neurosurgery, Universität Leipzig, Germany*

Abstract

The treatment of patients with pituitary adenoma requires the assessment of various patient data by the clinician. Because of their heterogeneity, they are stored in different sub-information systems, limiting a fast and easy access. The objective of this paper is to apply and test the tools provided by the openEHR Foundation to model the patient data relevant for diagnosis and treatment of the disease with the future intention to implement a centralised standard-based information platform. This platform should support the clinician in the treatment of the disease and improve the information exchange with other healthcare institutions. Some results of the domain modeling, so far obtained, are presented, and the advantages of openEHR emphasized. The free tools and the large database of existing structured and standard archetypes facilitated the modeling task. The separation of the domain modeling from the application development will support the next step of development of the information platform.

Keywords:
Patient data modeling, pituitary adenoma, archetypes, openEHR.

Introduction

A Pituitary adenoma is a kind of tumor that is located in the sella region of the brain. Usually, the tumor is benign, but it can affect the normal production of hormones, such as the sexual or growth hormone. These dysfunctions are responsible for example for infertility, erectile problems or enlargements of the extremities (hands, nose, ears) and of the internal organs (heart, kidneys) causing premature death. Because of the proximity of the pituitary gland with the optic chiasm, which is the region where the optic nerves cross, large pituitary tumors can moreover affect the vision of patients.

The diagnosis of the presence of a pituitary tumor is primarily performed based on MR imaging (figure 1). The position of the tumor in the pituitary gland and its size are estimated based on the image data. However, additive examinations, especially eye examinations, and blood tests which measure the quantity of hormones, provide complementary information about the kind of tumor. All these features strongly influence the choice of therapy [1].

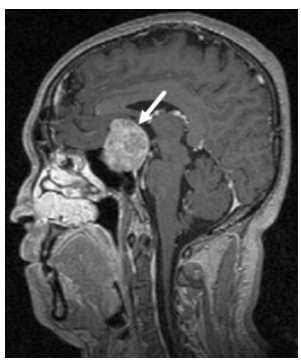


Figure 1 – MR data of patient revealing a large adenoma (white arrow) in the pituitary gland.

More generally, the treatment of diseases requires nowadays the acquisition and interpretation of a large amount of patient data of various formats, including image data, results of laboratory tests, findings or medical reports and letters. At hospitals, these patient data are digitalized in so-called electronic health records (EHR) and stored within a centralized database called hospital information systems (HIS) [2]. The history and all data of patients are therefore available at anytime for all clinicians within the hospital. However, the main drawbacks of these systems for the every-day use by clinicians are their heterogeneity and ,therefore ,their complexity of use.

Since patient data occurs in various data formats, different sub information systems constitute usually the HIS. Image data acquired within the hospital are stored in a Picture Archiving and Communication System (PACS) and are visualized and interpreted by the clinicians through a computer interface, a DICOM-viewer. Results of laboratory tests are also commonly managed within the specific software, specially designed for such particular data. External medical letters and findings are scanned and stored in specific databases accessible through their unique user interfaces. It becomes clear that it can be very complex for clinicians to find the relevant patient information among the available patient data distributed in the different software components of the HIS. The access to the patient data can moreover be very slow. Finally, the management and maintenance of such systems is very complex.

To overcome this situation we intend to implement a centralized and standard-based information platform dedicated to the treatment of patients with pituitary tumor according to the point of view of neurosurgeons. Domain modeling is a key step in such development [3]. It consists in representing and

storing the involved patient data in a suitable way. This paper focuses therefore on the modeling of the relevant patient data required for the treatment of the disease. The requirements associated with the domain modeling are presented. They form the basis to justify the choice of openEHR and associated tools. Finally, first results and experiences with the modeling are presented. In the future, the developed information platform should facilitate the diagnosis, therapy choice, treatment and monitoring of the disease by:

- Improving the access to patient data;
- Facilitating the interpretation of patient data;
- Increasing the exchange and re-use of patient data.

Materials and Methods

Selection of relevant patient data

The first step in the patient data modeling was the description of the clinical workflow for the treatment of patients with pituitary adenoma and the selection of relevant associated data. This was performed based on interviews with neurosurgeons.

Requirements for the patient data modeling

The second step consisted in representing the selected patient data according to models - to store and represent the information in the platform. This important task had to fulfill different requirements described in the following.

Definition of requirements. Content or fields of patient data models have to be determined. For example, the administrative data has to include personal information of patients, such as name, sex, birthday, address, identifier number and entry date.

Standard representation. The patient data has to be represented in a way so that it is understable to all actors. For example dates and time points need to be uniquely represented since they are differently recorded according to continents and countries [4].

Structured representation. In the context of the treatment of pituitary adenomas, a neurosurgeon makes the decision for the treatment based on images, laboratory data or information from the clinical reports provided by other clinical specialties. An unexperimented neurosurgeon is little familiar with these patient data. Structuring the information by enhancing important facts or conclusions facilitates the interpretation [5].

Interoperability. The use of existing norms and standards for modeling is a key point to improve the exchange of patient data between healthcare institutions [6].

Support for the implementation of the information platform. Patient data modeling forms the basis for the implementation of an information platform and most widely support and facilitate it [3].

openEHR

The openEHR Foundation is described on its web site (http://www.openehr.org/) as follows: “OpenEHR is a virtual community working on interoperability and computability in e-health. Its main focus is electronic patient records (EHRs) and systems.” [7, 8]. Different applications were already developed based on openEHR [9, 10], for example in the field of neonatology [11] and epidemiological studies [12]. The tools provided by openEHR have important advantages for the development of our application.

Free availability. The philosophy of openEHR is to share the knowledge in the field of patient data modeling. Several tools are available to all for free. Users are encouraged to reuse these archetypes for their applications. These archetypes are good templates for the development of new archetypes. The database is totally free available so that each one can contribute on its enlargement. A discussion forum enables support and collaboration.

Standard-based archetype database. Patient data models are called archetypes in openEHR. A large database of existing archetypes is available. They were developed based on rules and structures, now recognized as standard. For example, a large set of archetypes represents the results or findings of a medical examination. A main component of these archetypes are conclusions of the examination -that are crucial for the diagnosis of the disease. However, further information about the requester of the examination, the name of the healthcare institutions where the examination was performed as well as the data and time are also important and included in the templates in general. Moreover, it is possible to link the archetypes with existing terminologies, for example, to code some text like the modality of image data.

Archetype Editor tool. This editor aims at supplementing existing archetypes for a specific applications, at building new archetypes and at assembling archetypes for the development of an application.

Separation of domain modeling and application development. The idea of openEHR is to separate the domain modeling from the application development. Therefore, the modeling of patient data can be performed by physicians or noncomputer scientists. The tools proposed by openEHR, like the *Archetype Editor*, are intended to be easy to use. A second software tool, the *Template Designer*, was developed to join all archetypes relevant for one application. The existing archetypes include a large number of information fields -that are not all relevant for a given application. With the *Template Designer,* it is possible to define constraints on the archetypes, for example the occurrence of an archetype. Some information fields can be prohibited. An XML-document is then generated from the domain modeling, on which the computer development of the information platform is based.

We tested and applied the tools provided by openEHR for the domain modeling of the treatment of pituitary adenomas.

Results

Clinical workflow and patient data set

The clinical workflow of treatment of patients with pituitary adenoma includes mainly three phases: firstly diagnosis of the disease, secondly therapy choice and treatment, and thirdly control of treatment and monitoring of the disease (figure 2).

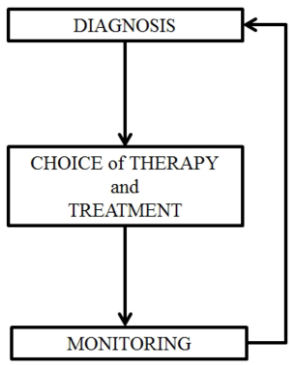


Figure 2 – Clinical workflow of the treatment of pituitary adenomas.

Diagnosis phase. During the first visit of the patient at the hospital, the administrative data of the patient are recorded. Diagnosis of the disease is primarily performed based on MR data which reveals the possible presence of a tumor and, if it is positive, provides an estimation of its size and position. Moreover, results of laboratory tests, eye examinations (more specifically the measurement of visual acuity and visual field) and clinical symptoms expressed by patients strongly influence the choice of therapy. The most common options are pharmacotherapy, surgery, observation of the disease development and combinations of them.

Treatment phase. In our application, the patient data are collected only in the pre-, intra- and postoperative phase of the surgical therapy. At the University Hospital of Leipzig transsphenoidal endoscopic operations is performed. An Endoscope and surgical instruments are inserted through the nose and the sphenoid bone till the sphenoidal sinus cavity - - to remove the tumor. The operation is assisted using a navigation system that indicates at anytime the position of the surgical instruments on CT and MR image data acquired one day before. Bone structures are well revealed on the CT image to guide the surgeon on the way to the tumor, while the MR data enhance the soft tissue, like the pituitary adenoma, in the brain area. During the operation planning, the surgeon can semi-automatically segment the tumor, especially when it is badly visible in the images. In the operating room, endoscopic video images are generated and stored for operation documentation. Postoperatively, a CT or MR image is generated in order to reveal possible bleeding or to detect possible remnants of tumor. Finally, the surgeon draws up the operation progress on a medical report.

Monitoring phase. Patient data involved in the monitoring of the treatment and of the disease are the same than those used for the diagnosis.

In conclusion, Table 1 lists patient data involved in the different clinical phases and the relevant information they provide to the neurosurgeon.

Table 1 – Patient data involved in the diagnosis, surgical treatment and monitoring of pituitary adenomas

Phase	Patient data	Relevant information
1	Master data	PID / Sex / Age
	MR data of the head	Tumor size and location
	Clinical symptoms	List of symptoms
	Laboratory tests	Hormone concentration in blood above or under normal values
	Eye examination	Measurements of the visual acuity and visual field
2	Preop. MR data	Tumor size and position (segmentation) for surgery planning
	Preop. CT data	Bone structures (visualization) for surgery planning
	Intraop. endoscopic video	Selected images for documentation
	Operation report	Documentation
	Postop. CT or MR	Presence of bleeding / Detection of remnants of tumor
3	Identical than in phase 1	

preop.: preoperative
intraop.: intraoperative
postop.: postoperative

Clinicians pointed out that four kinds of information must be attached to each archetype:

- Original text or image data
- Healthcare institution which performed the medical examination
- Date and time of the examination
- Comments of the neurosurgeon

Domain modeling with openEHR

The patient data listed in Table 1 were modeled using seven archetypes from the openEHR database and five self-implemented archetypes (Table 2). Since imaging examination are required for the three phases of patient treatment, its modeling is described in more detail in the following.

Table 2 – List of archetypes used for pituitary adenoma domain modeling

Patient data	Archetype name	OpenEHR	New
Master data	*Research study*		x
Image data	*Imaging examination results*	x	
Radiological findings	*Organisation*	x	
	Pituitary adenoma classification		x
Clinical symptoms	*Pituitary adenoma clinical symptoms*		x
Laboratory tests	*Pituitary hormones*		x
Eye examination	*Visual acuity*	x	
	Visual field measurement	x	
Choice of therapy	*Procedure request*	x	
Intraoperative image data Operation report	*Procedure undertaken*	x	
Histological findings	*Histopathology*	x	
	Pituitary adenoma histology		x

Example: imaging examination modeling

Image data and radiological findings

The existing archetype *imaging examination result* from the openEHR database, models image data with the corresponding radiological finding. The finding is structured in different sub-parts including the examination name, the examined anatomical site, the date, time and status of the given result, the radiological diagnosis and possible conclusion or comment. A link to the original text data can be specified. Moreover, the imaging examination can be documented: who requested and performed the image acquisition, when was the acquisition performed, what was the patient position. A reference image of a current view can be attached.

Usually, only selected fields are required according to the medical application. In agreement with the neurosurgeon the following fields were kept:

- "Imaging modality": possible choices were coded;
- "Conclusion": in this field can the neurosurgeon add his comments;
- "Examination result representation" includes the link to the original radiological report;
- "Name of organisation" which performed the imaging examination;
- "Image details" includes the date and time of image acquisition as well as a reference image, possibly with the tumor segmentation.

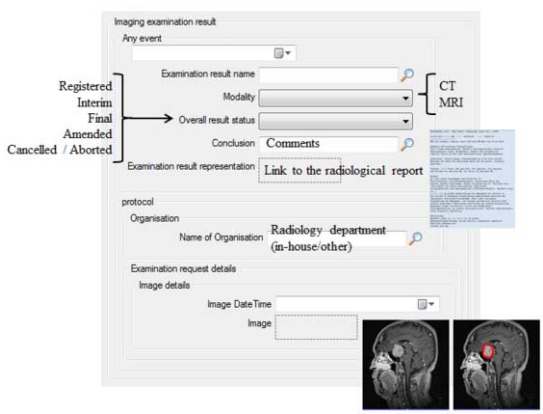


Figure 3 – Preview performed with the tool "TemplateDesigner" of the selected fields of the existing archetype "Imaging examination result".

The archetype was improved using the tool *Archetype Editor*. The choices for the field "Imaging modality" were coded using an internal code because only CT or MR data is involved. The field "Name of organisation" was linked to the existing archetype *Organisation*, which reports details of an organisation or agency.

In the tool *Template Designer*, the other fields not required for our application were set as prohibited so that they will not appear in the information platform. Moreover, the fields "Examination result name" and "Overall result status" are mandatory by definition in the archetype. A preview of the mask corresponding to the modified archetype can be generated and is shown in figure 3.

Pituitary adenoma classification

Based on the observation of the image data and on the reading of the radiological findings, the neurosurgeon infers a classification of the pituitary according to its size and location (Table 3). Since this classification is specific to our application, we developed ourself the corresponding archetype.

Table 3 – Description of the new archetype "pituitary adenoma classification"

Fields	Description	Modeling
Number of tumors	Between 1 and 4	Number
Maximum size	Maximum tumor size estimated in the three radiological directions	Quantity
Macroadenoma	Tumor size ≥ 10 mm	Boolean
Microadenoma	Tumor size ≤ 10 mm	Boolean
Localization	Tumor localization and link with adjacent structures	Boolean
Symmetrical growth	Performed in the coronal view	Boolean
Hardy classification	Classification based on tumor size and localization: Grade 0 to Grade IV	Coded text

The new archetype was developed using the tool *Archetype Editor*. Different kinds of data can be included into the archetype like text, numbers or boolean values. The unit can be attached to a data "quantity", like it is the case for the tumor size given in millimeter. It is possible to specify the range of value which are entered by the user. For example, the number of tumors was limited to four, the values for maximum tumor size must be positive. The Hardy classification [13] was coded using an internal code from Grade 0 until Grade IV. A preview of the archetype is shown in figure 4.

Discussion and conclusion

Patient data modeling is a crucial step towards the development of information systems in health care domain. The openEHR Foundation proposes free software that is well adapted to support this task. The domain modeling is totally separated from the application development. This enables a medical expert with limited experience in computer development to implement the respective domain knowledge. The large database of archetypes provided by openEHR was developed according to standards. The tool *Archetype Editor* enables to complement existing archetypes, to implement new archetypes and to specify constraints on the archetypes. Then, they are put together in the tool *Template Designer*. A preview of the web mask corresponding to the domain modeling can be generated. The model can be exported as an XML file which forms the basis for the development of the web mask. The standardisation of the models of patient data should encourage the exchange between healthcare organisations. In general, the use of the openEHR tools require experience since a documentation is available only to a limited extend or difficult to find. This can be a limitation for clinicians.

So far we implemented a prototype of the information platform based on a couple of patient data. This proof of concept project showed us the feasibility of developing an information system based on openEHR. The future objective

of this work is the development of therapeutic assistance system to support clinicians in decision making for the treatment of patients with pituitary adenomas. This requires the extraction of semantic information from patient data, whose representation in archetypes is supported by openEHR. For example adverse events concerning the patient or occurring during the treatment can strongly influence therapeutic decisions. Specific new text mining and image processing tools have to be developed in research environment and included within our information platform [14, 15].

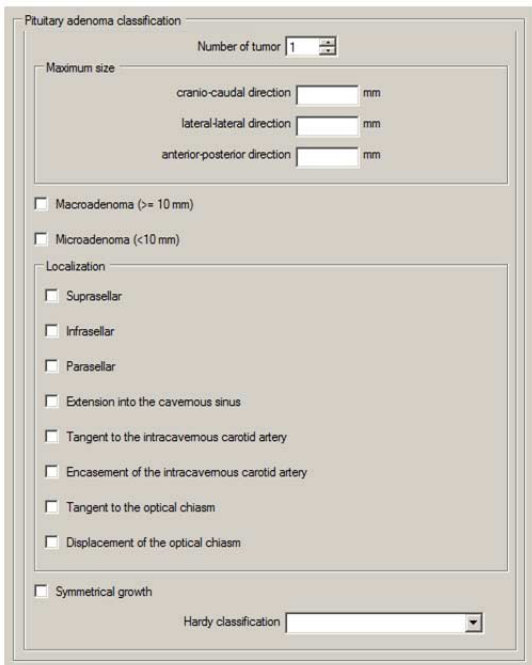


Figure 4 – Preview of the new archetype "Pituitary adenoma classification" recording the tumor classification performed based on image patient data (mask generated by the "Template Designer").

References

[1] Scanarini M. Mingrino S. Functional classification of pituitary adenomas. Acta Neurochirurgica 1980: 52(3–4): 195-202.

[2] Bisbal J, Berry D. An analysis framework for electronic health record systems. Methods Inf Med 2011: 50(2): 180-189.

[3] Atalag K, Yang HY. From openEHR Domain Models to Advanced User Interfaces: A Case Study in Endoscopy. In: Health Informatics New Zealand Conference 2010.

[4] Beale T. Health information standards manisfesto. http://workflow.healthbase.info/monographs/Tom_Beale_his_manifesto.pdf 2001.

[5] Denecke K. Model-based Decision Support: Requirements and Future for its Application in Surgery. Biomedical Engineering / Biomedizinische Technik 2013: 58(S1).

[6] Costa CM, Menarguez-Tortosa M, Fernandez-Breis JT. Clinical data interoperability based on archetype transformation. J Biomed Inform 2011: 44(5): 869-880.

[7] Beale T, Heard S. openEHR Architecture Overview. http://www.openehr.org/releases/1.0.2/architecture/overview.pdf 2008.

[8] Garde S, Knaup P, Hovenga EJS, Heard S. Towards semantic interoperability for electronic health records. Methods Inf Med 2007: 46: 332-343.

[9] Kashfi H, Torgersson O. A migration to an openEHR-based clinical application. Stud Health Technol Inform. 2009: 150: 152-56.

[10] Mutiara AB, Muslim A, Oswari T, Miharja RA. A Model of OpenEHR Based Electronic Medical Record In Indonesia. European Journal of Scientific Research (EJSR) 2012: 90(3).

[11] Buck J, Garde S, Kohl CD, Knaup-Gregori P. Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach. Int J Med Inf 2009: 78(8): 521-31.

[12] Freire SM, Sundvall E, Karlsson D, Lambrix P. Performance of XML Databases for Epidemiological Queries in Archetype-Based EHRs. Scand Conf Health Inform 2012: 13: 51-57.

[13] Hardy J, Vezinaj L. Transsphenoidal neurosurgery of intracranial neoplasm. Adv Neurol 1976: 15: 261-273.

[14] Denecke K, Bernauer J. Extracting Specific Medical Data Using Semantic Structures. AIME 2007: 4594: 257-264.

[15] Kang N, Singh B, Bui C, Afzal Z, van Mulligen EM, Kors JA. Knowledge-based extraction of adverse drug events from biomedical text. BMC Bioinformatics 2014: 4: 15:64.

Address for correspondence

Claire Chalopin / Innovation Center Computer Assisted Surgery
Semmelweisstr. 14
D - 04103 Leipzig, Germany

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-183

Evaluating Business Value of IT in Healthcare: Three Clinical Practices from Australia and the US

Peter Haddad [a,c], Jonathan L. Schaffer[b], Nilmini Wickramasinghe [a,c]

[a] *RMIT University Victoria, Australia*
[b] *Cleveland Clinic, Ohio, USA*
[c] *Epworth HealthCare, Victoria, Australia*

Abstract

Exponentially increasing costs in healthcare coupled with poor quality and limited access have motivated the healthcare industry to turn to IS/IT solutions to overcome these issues and facilitate superior healthcare delivery. In an environment of rapid development of new clinical informatics solutions claiming to provide better healthcare delivery, there is a paucity of systematic frameworks to robustly measure the actual value of these systems. The promised business value of these solutions has been elusive; hence, this study offers an approach for the evaluation of the business value of health IS/IT solutions based on a conceptual model, which has been validated using three clinical case studies.

Keywords:

Healthcare, Information Techonology, Information Systems, Business Value of IT, Computerized Systems.

Introduction

Healthcare spending is increasing exponentially throughout the world [1], necessitating decision makers to take action to address this dilema[2,3]. Not only are the expenses creating pressure for the healthcare industry where resources are already strained, but also increasing problems relating to heathcare quality and patient safety. According to a report by the Institute of Medicine [2], more than 98 000 Americans die annually, and more than ten times that number are injured as a result of this problematic situation. Furthermore, these figures are projected to be getting even worse. A 2010 study[3] estimates that ~180 000 Americans will die per year, and another 2013 study [4] estimates this to be between 210 000 and 440 000 patients. Clearly, this is a serious and untenable situation, not to mention the tremendous cost of all of this for the American healthcare system.

Investing in information technology has been touted as a new strategy to correct all of these issues and achieve the desired cost/quality balance (input/outcome) [5, 6]. As a result, OECD countries are now heavily investing in health information systems [7, 1]. This growth has been driven to a large extent by success stories for such investments in other sectors of the service industry such as retail, education and logistics [8]. While most healthcare providers globally have started investing in health IT, we are yet to witness the promised increased business value from these investments [8]. However, at present, too many significant failures with health IT investments permeate the literature [9]. Thus, understanding how information technology can help generate business value in healthcare becomes a priority.

The impact of information technology on the organisational performance (i.e business value of IT) has been a major theme in the research since 1992 [10]. Since this time, researchers have not stopped investigating this area [11, 12]. The scope of these studies, however, has been limited to retail and manufacturing industries, and has not covered the healthcare industry. The reason behind this could be due to the complexity of healthcare delivery, consisting of many systems and subsystems; as all of these as ecosystems must be considered if a true sense of value is to be established [13, 8].

Our exploratory study aims to examine the business value of IT in healthcare. To do so, we propose a comprehensive conceptual model that covers almost all types of IS/IT, and all levels of healthcare delivery. This model is then tested against three exemplar case studies; noted by Yin to be a prudent methodology under these conditions [14].

Business Value of IT in Healthcare: A Conceptual Model

Although the business value of IT has been previously investigated, the current literature still has several key deficits, notably a lack of understanding of the relationship between IT investments and firm performance [12, 8]. Many researchers believe that an appropriate strategy is to move from the question of whether IT creates value to how, when and why benefits occur or fail to do so [15]. To develop a conceptual model to support this line of inquiry, we look at IS/IT investments through the lens of their management objectives. This lens is derived from the work of Peter Weill, who classified IT investments into four types: informational, transactional, infrastructure, and strategic systems; which forms what is called 'IT portfolio', and in turn represents an organization's entire investment in IT [16].

Classifying IT investments in this way forms the first step toward developing a conceptual model. The second step is recognising the socio-technical aspects, a differentiating feature which represents one of the big differences between healthcare and other service industries [17]. According to this conceptualisation, healthcare delivery can be divided into four interrelated layers: healthcare ecosystem (society), system structure (organisation), delivery operations (processes), and clinical practices (people). Further, the efficiencies that can be achieved at the lowest level (clinical practices) are limited by the nature of the next level (delivery operations).

The conceptual model suggested by this research aims at reflecting the complexity of healthcare delivery and the vast range of IS/IT investments in this industry. This is done by taking both the IT Portfolio [16] and the Healthcare Delivery

Enterprise Model [17] into consideration as depicted in Figure 1.

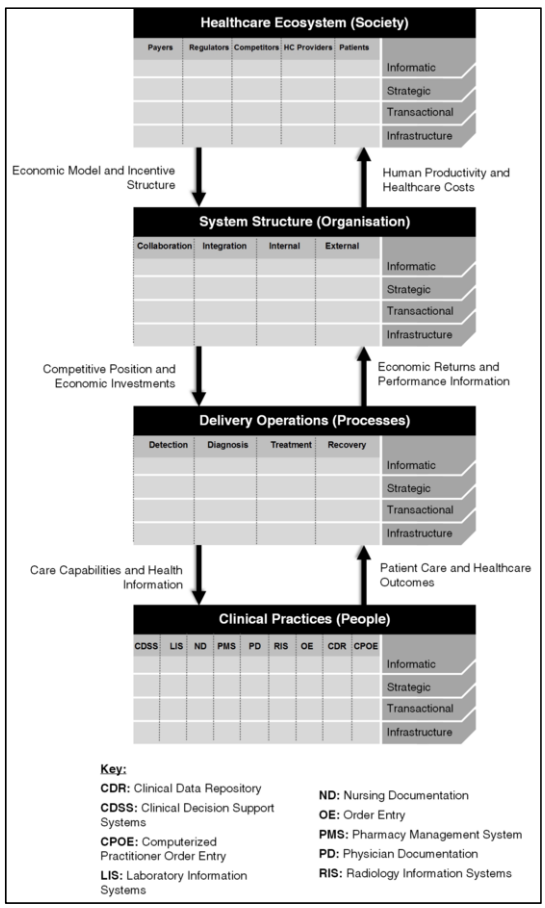


Figure 1: Business Value of IT in Healthcare: A Conceptual Model

Methods

A case study is the appropriate method to illustrate critical issues or key points in such a context [14]. For the purpose of this paper, we use three exemplary case vignettes that highlight three clinical practices in the conceptual model. The first is an Australian nursing documentation (ND) system, the second is an American computerised practitioner order entry (CPOE) system, and the third is an online second medical opinion application (MyConsult). For ND and CPOE systems, data collection took place between 2013 and 2014 in Australia. The methods used were mainly a series of 28 semi-structured interviews with the main stackholders at a leading private healthcare provider in Australia. This includes professionals from three groups: clinicals, IT and business people. All of the interviews were transcribed and qualitatively analysed using Nvivo 10. MyConsult has a 12 year history of adminstrative and clinical data collection and we focused on the last three years examining the clinical outcomes of the second opinion reports combined with administrative data. Additionally, semistructured interviews were also completed as noted with the ND and CPOE groups. These interviews were conducted with 12 department and institute chairs or their respective administrators as well as 8 members of the Distance Health and Information Technology Division and Finance Department staff at the Cleveland Clinic.

The use of three significantly different systems is intentional. The aim of having these three different systems is to test the robustness of our conceptual model and thus to be used to evaluate the business value of IT in different and distinct contexts and circumstances. Testing the model with similar cases will only serve to demonstrate the usefulness of the model in a narrow application.

We look at different IS/IT solutions in two different countries whose healthcare ecosystems are different. This enables us to test whether or not any amendments on the proposed conceptual model are needed.

MyConsult is a comprehensive online application that provides an online model for a patient to request a second opinion from experts at the Cleveland Clinic. As the only US institution offering such a comprehensive model that is secure, patient specific and incorporated into clinical practice, MyConsult offers patients, physicians and families the opportunities to remove the geographic barriers to care and receive a formal report through the online secure portal. While the consult is real, it is the visit that is virtual (please note that Ask MayoExpert and WebMD systems were not considered comparable). The MyConsult model provides an excellent example of a novel technological approach to clinical care in a specific regulatory environment using an application that has been custom designed and developed by the technical teams at the Cleveland Clinic.

MyConsult pricing, market and program research indicates that those requesting MyConsult were in fact looking for specific program attributes, a comprehensive review of general and diagnosis specific questionnaires and the data set that comprised the first opinion; all resulting in a formal written report that is shared with the patient's treating physician team in their home market. The development structure permits MyConsult to quantify the business value of the development work in conjunction with the clinical value received by patients.

All who were interviewed noted that the clinic's online second opinion system was far superior than any other system they had used or come across.

Case vignettes

Case Vignette 1: Nursing Documentation System

For the purpose of this paper, we examined a nursing documentation system that was designed and built in Australia. The developers envisage that this system will replace paper-based nursing documentation in hospital acute care wards in both private and public hospitals. The objective of this system is to allocate more time for patient care by putting smart terminals at the patient bedside, which would create more interaction between nurses and patients. The system is still under research, where a research team consisting of both nursing and IS researchers are examining the system, and providing feedback to the designers, who are, in turn, releasing new versions constantly. By making these smart terminals available by the bedside, a higher level of collaboration between ward nurses and other healthcare professionals is expected; according to the designers of this system.

Case Vignette 2: Computerised Practitioner Order Entry (CPOE) System

The studied CPOE system in this paper is part of a comprehensive information and image management system

specialised in oncology care. The vendor offers three modules of this system: radiation, medical, and surgical. In our case study, the two purchased modules are radiation and medical, while the surgical module had not been acquired.

The aim of implementing this system is to replace the manual scheduling of pateients by electronic means, which would help make the processes of admission, treatment, and post treatment care plans smoother and with less errors or delay.

The radiation module has been implemented in our case study for two years now, while the medical module has just been acquired. Globally, the system has been in use for 10-15 years. In the Australian context, it is been in use for about 7 years.

Case Vignette 3: Online Second Medical Opinion Application

The Cleveland Clinic MyConsult application removes the geographic barriers to care by providing an online second medical opinion for those diagnoses with sufficient objective data upon which a remote evaluation can be completed. The focus is to determine if the diagnosis is correct and if the patient is receiving the optimal treatment. These patients can receive a second opinion without leaving their homes. Integration with the existing IT infrastructure includes network, interface engine, LDAP, DB, upload, EMR, Security, Financial, and Materials Management. Three structural layers classified the business value discussion: first the external system to receive and organize these requests, second, the internal system to coordinate, consolidate, and organize per diagnosis, and third, geographic issues from the international, national and regional unique markets.

Results

The participants for ND and CPOE had the following demographic patterns: 46.42% of the 28 interviewees were females and 53.50% males. In terms of specialty, the interviewees were clinicians (nurses, oncologists and clinical auditors) (26.67%), IT personnel (33.33%), and executives (including clinical executives) (40%). All the interviewees had more than 15 years' experience and 89% of them had more than 7 years at the case study. Majority of the interviewees were within the age group 41-50 years (46.7%), 51-60 years (26.7%), 31-40 (13.3%), and 61-70 (13.3%).

For MyConsult, 80% of the participants were male and were comprised of 40% clinical staff and 60% administrative staff. All of the administrative staff had more than five years of experience while the clinical staff had more than 20 years of experience.

The Role of IT in Healthcare

A number of factors that help generate business value out of IT were identified during the interviews. The majority of the interviewed staff recognized the role IT can play to enhance the organisational performance of healthcare providers. From a nursing perspective, IT *"could absolutely and should facilitate the patient journey through health, generally, from the community into the acute care environment and then back into the community. Certainly, that provides the quality of continuity, but I think it also provides a safer environment for that patient"* [NAS#1]. From another clinical point of view, IT is an important part of contemporary healthcare organisations *"in that it measures what it can be measured and more importantly it facilitates processes"* [CAS#3]. Facilitating processes is not limited to clinical processes, but it also covers the business side of healthcare provision. *"not only does it reduce the cost or increase the efficiency of processes, but also it enhances the integration with other partners like the health care funds".* [CIO].

Business Value of IT: Problems and Requirements

In order for IT to generate business value there are a number of organisational and conceptual requirements that must be met. The big differenece between health organisations and other organisations is that *"healthcare in general, and hospitals in particular are not organizations in the traditional sense. They don't have hierarchical line authority from top to bottom, particularly in the private sector. They're not organizations. They're loosely affiliated groups of people trying to do the same thing".* [CAS#2]. Thus, the business value of clinical IT is hard to be montized. Asking about the reasons behind this, the interviewee suggested that the apprenticeship system through which the doctors are raised up is to be blamed: "*Generally in that sort of training, process is not a reflection. What I would call a deep reflection of what is the nature of your work process as you get with the apprenticeship training*". As a solution to this conceptual issue, the Director Clinical Order and the Group Manager Medical Services suggests that "*I suppose if we train doctors to think of themselves as part of a system rather than as independent practitioners, that would achieve that end*".

The other problem with IT investments, and especially the clinical IT systems, is that healthcare organisations tend to be solution-focused, more than problem-focused. "*so they come in solutions and then look for the problem that it's solving as opposed to what problem do I need, what actual things do I need to resolve*" [PMO#1].

Case Vignette 1: Nursing Documentation System

The system has shown high fidelity to the processes of nursing care delivery in the sense of time consumed for nursing documentation. This time fell from 15.7% to 6.4% after the inroduction of this system, freeing up over 9% of nurse time from record-keeping. This decrease has allowed an increase in the proportion of nursing activities devoted to patients assessment from 5.2% to 9.1%. The way the system was designed (smart terminals by the bedside) has increased transparency in healthcare provision. This has been achieved by increasing the nursing activities performed by the bedside by up to 48.1% since the introduction of this system, which has created better platforms for longer interaction between the nurses and the patients (23.6% of time nurses spend with patients compared to 7.95% before the system). [1]

Although maximsing the care time and minimizing the time dedicated to other nursing activities is important, it is "*much more complex than just taking some aspects of the role away"* [NAS#1], citing the the big strike back in 1980's in Victoria, Australia around non-nursing duties. Back then, nurses were continually doing duties that were not direct patient care, like some cleaning activities, supervising things that probably should have been done by housekeeping, and lots of administrative paperwork. Although a lot of those responsibilities were taken away from nursing, there has not been an evaluation that shows this has enhanced nursing care. *"I think that even with our new technology that we're testing now, nurses find other things to do"* [NAS#4].

During the interviews, our interviewees gave us the impressions that the concentration should be focused on the quality of time spent with patients not the duration of it. This is complex though, and requires a "*whole of care redesign. IT has the ability to facilitate this, but there are human factors should not be left behind*". [NAS#3].

[1] These results are the summary of a research project to evaluate the fidelity and usability of the ND system. The reference is available upon request.

Another important aspect of introducing state-of-the-art IT systems into healthcare, which is visible to the patients is the reputation, which is a *"fundamental part of the business"* [PMO#5]. Consequently, *"if patients feel that even just on face value, that the nurses are not scribbling on a piece of paper, but actually, they've got an IT system that provides confidence in the organization".* [NAS#3].

Two other problems were identified with this system. First, it is limited to ward care, and it does not fit well within the scheme of care in the hospital. *"I'm a bit concerned that you've got an end-to-end process where a patient comes in. They're booked. They're pre-admitted. They're admitted. They have an inpatient stay. In that inpatient stay, there's a section in theatre. They're then discharged and then go home, whatever. This system may be a part of that whole process. It won't do the theatre piece so when the patient goes into theatre"* [CIO]. Thus, The fit of any IS/IT system into the care strategy of any healthcare provider needs to be clearly articulated.

Second, the system is still under development, and "*health care companies don't tend to like to be at the bleeding edge*" [CIO], as "*we are healthcare business, not an IT shop*" [CCIO].

Case vignette 2: Computerised Practitioner Order Entry (CPOE) System

The studied CPOE system in this paper is dedicated to cancer care. The difference between the drug prescribing systems in cancer care and other systems in different care contexts is its complexity. "*Drug prescribing systems typically prescribe a single drug for blood pressure or antibiotics and dizzy dotted period. In the cancer treatment, there are things called chemotherapy private calls, which are often five or six drugs given at very specific individuals, specific doses for lengthy period and the prescribing function in those is quite complex because they have to be given according to certain recognized protocols".* [OAS#1]. The studied COPE system provides these complex funtions.

The introduction of this system has positively impacted the patients'outcomes. *"There's been a significant reduction in errors because we prescribe according to the protocol. These protocols are quite common and human error in transcribing is common. It's reduced legibility errors, all that sort of stuff."*[OAS#2].

The problems with implementing this specific system can be classified into two categories: the system characteristics, and the organisational characteristics.

The System Characteristics

The system is an American model. *"it is an international software that struggles with local billing practices which are different everywhere, and so are the drug protocols. It needs more domestication because it's an American product"* [OAS#1].

Second, the studied hospital has two of the three available modules of this product. These two modules were be described as: *"one really is good, the other is okay and we are trying to get them to talk to each other."* [PMO#2].

The Organisational Characteristics

The studied hospital is a private hospital, where the majority of the oncologists are visitor medical officers (VMOs). This group of knowledge workers have invested heavily (both emotionally and financially) in their own IT systems. Although there's a lot of mutual interest in IT between the VMOs and the hospital, "*they're talking about different sorts of IT. That leads to a massive amount of misunderstanding*" [OAS#1]. This misunderstanding seems to result in incompatibility between the two parties. This incompatibility is more conceptual rather than technical: *"The incompatibility between IT systems is fixable and it's mostly achieved by standards like HL7. But the incompatibility that I'm talking about between medical professional staff in hospital is a conceptually incompatibility"* [OAS#2].

The other organisational problem is that IT in this hospital is mainly controlled by a project management approach, where project managers, who report to "project sponsors", have other full day-to-day responsibilties. This results in delays in meeting the milestones, and less tangible impact of the IT system in the short-term.

Case vignette 3: Online Second Medical Opinion Application

The MyConsult application has provided second opinions to patients from over 105 countries for the past 12 years. While most of the users are consumers who locate the service through traditional search services, over 13.5 millions are provided access to MyConsult as a health care benefit from their employer. Evaluating the data streams and interviews focused on the business, administrative technology, and clinical inputs and outputs. MyConsult improves efficiency for patients, their doctors and consultants; empowers and engages patients with personalized information and the productivity enhancements substantiates the value produced in the program.

Each of the internal and external costs were evaluated as depicted in Table 1. Furthermore, these costs were classified based on the level of healthcare delivery.

Table 1- Mapping MyConsult to the Conceptual Model

Internal and External Costs	MyConsult Operational Response	Parameter	IT Layers
Working the request in individual offices	Centralized, standardized MyConsult Clinical Operations Center	Clinical Practice - Internal variability in case labor becomes standardized	Transactional Infrastructure
Customer service distributed	Trained, highly experienced team	Delivery Operations - Patient experience	Strategic Infrastructure
Physician participation and expertise matching	Institute derived participation with clinical leaders	System structure - Internal physician satisfaction	Strategic Transactional Infrastructure
Variable response format	Standardize report format	Delivery operations - Patient experience	Transactional
Readability and comprehension variable	Doctor translated into patient English	Delivery operations - Patient experience	Informatic Transactional
Conflict with other duties	Incorporated in duties	System structure, Clinical practices - Internal physician satisfaction	System structure Delivery operations Clinical Practices
Expectations from patient not always clear	General and diagnosis specific questionnaires	Delivery operations, patient satisfaction	Healthcare ecosystem Informatic
Medical record collection challenges	Offering medical record collection services	System structure, Delivery operations - Patient experience	Informatic, Strategic, Transactional, Infrastructure

Overall, the enterprise adoption and adaptation of the application was supported by the clinical leadership driving participation and diagnosis coverage, streamlining and optimizing online and clinical processes, and integrating into clinical practices and schedules. For patients, the concurrence of diagnosis approached 75% while the concurrence of treatment recommendations approached only 38%, further demonstrating value provided to patients.

Discussion

In healthcare, demonstrating business value of IT is challenging. This is largely due to the complex nature of healthcare operations, the web of players that must interact with any particular IT solution, as well as structural and organizational issues. An integrative model to assess the value of a clincial informatics solution requires consideration of the key stakeholder's perspectives of healthcare value. The use of the IT portfolio classificiations of infrastructure, transactional, informational, and strategic IT provides an excellent basis for organizing these perspectivies [16]. When combined with the

Enterprise of Healthcare Delivery Model classifications of clinical practices (people); (ii) delivery operations (processes); (iii) system structure (organizations); and (iv) healthcare ecosystem (society) [17], this model becomes a powerful tool for assessing the business value of healthcare IT.

The three case studies used to illustrate and demonstrate the fidelity of the proposed solution are taken from clinical contexts that exhibit these challenges. For example, unique structural considerations at the Cleveland Clinic include that its medical staff are salaried, while unique structural considerations for private hospitals in Australia include that their doctors are typically visiting medical officers (VMOs) and not salaried. Such structural differences severely affect the effectiveness of enterprise wide clinical IT systems such as the ones discussed in the respective scenarios, and thus also the ultimate business value of the IT that can be realized. While still under development, the proposed assessment tool is both flexible and rich enough to provide useful information to executives in all these scenarios. We believe that this is a key strength of the tool. At this stage, assessments are more qualitative than quantitative, but we anticipate developing quantitative scoring systems in the future. What is particularly attractive about the proposed tool is the generalized applicability across all healthcare IT solutions, which means that now it will be possible to carefully examine and compare relative values of different IT solutions in healthcare contexts; something which to date has been extremely difficult, if not impossible, to do.

Conclusion

As the adoption, assimilation and diffusion of IT in healthcare increases coupled with the pressure for healthcare organizations to demonstrate improvements with respect to access, quality and value for patients; the need to objectively assess the value afforded by IT investments will increase. Our research serves to begin to address this need through the development of an integrative value of IT assessment tool and has demonstrated its validity by illustrating its use in three different clinical contexts. There is a clear need for a systematic, robust framework to assess the business value of these clinical informatics solutions. Future research will focus on developing the framework further and continually testing it with the many IT solutions that are becoming available to assist in directing appropriate resources.

Conflict of Interest

There is no conflict of interest by any member of the study team or from the participants in this research.

Address for correspondence

peter.haddad@rmit.edu.au

References

[1] OECD. Strengthening Health Information Infrastructure for Health Care Quality Governance. OECD Publishing; 2013.

[2] Medicine Io. BUILDING A BETTER DELIVERY SYSTEM: A New Engineering/Health Care Partnership: Institute of Medicine2005.

[3] DR L. Adverse Events in Hospitals: National Incidence Among Medicare Beneficiaries. Washington, DC: US Department of Health and Human Services, Office of the Inspector General2010.

[4] James JT. A new, evidence-based estimate of patient harms associated with hospital care. p. 122-8.

[5] Berner ES. Implementation Challenges for Clinical and Research Information Systems: Recommendations from the 2007 Winter Symposium of the American College of Medical Informatics. Journal of the American Medical Informatics Association.15(3):281-2. doi:10.1197/jamia.M2581.

[6] Kilbridge PM, Classen DC. The informatics opportunities at the intersection of patient safety and clinical informatics. Journal of the American Medical Informatics Association.15(4):397. doi:10.1197/jamia.M2735.

[7] Wickramasinghe N, Schaffer J. Realizing Value Driven e-Health Solutions. Washington DC: IBM Center for the Business of Government2010.

[8] Peter Haddad, Mark Gregory, Nilmini Wickramasinghe. Evaluating Business Value of IT in Healthcare in Australia: The Case of an Intelligent Operational Planning Support Tool Solution. Bled 2014: AIS Electronic Library; 2014.

[9] Kaplan B, Harris-Salamone KD. Health IT Success and Failure: Recommendations from Literature and an AMIA Workshop. J Am Med Inf Assoc. 2009;16(3):291-9. doi:10.1197/jamia.M2997.

[10] Weill P. The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. Information Systems Research. 1992;3(4):307-33.

[11] Silvius AJG. The business value of IT: A conceptual model for understanding and selecting valuation methods. 2011.

[12] Masli A, Richardson VJ, Sanchez JM, Smith RE. The business value of IT: A synthesis and framework of archival research. Journal of Information Systems. 2011;25(2):81-116. doi:10.2308/isys-10117.

[13] von Lubitz D, Patricelli F. Network-centric healthcare operations: Data warehousing and the associated telecommunications platforms. International Journal of Services and Standards. 2007;3(1):97-119. doi:10.1504/IJSS.2007.011830.

[14] Yin RKa. Case study research : design and methods. Fifth edition. ed. SAGE; 2014.

[15] Melville N, Kraemer K, Gurbaxani V. Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. MIS Quarterly. 2004;28(2):283-322.

[16] Weill P, Broadbent M. Leveraging the new infrastructure : how market leaders capitalize on information technology. Boston, Mass.: Harvard Business School Press; 1998.

[17] Rouse WB, Cortese DA. Engineering the system of healthcare delivery. Amsterdam: Amsterdam: IOS Press; 2010.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-188

Toward User-Centered Patient Safety Event Reporting System: A Trial of Text Prediction in Clinical Data Entry

Lei Hua[ab], Yang Gong [a]

[a] *School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX*
[b] *Informatics Institute, University of Missouri, Columbia, MO*

Abstract

As a primary source for learning from lessons, patient safety event reporting systems have been widely adopted. Nevertheless, underreporting and low quality of reports pervade the system. To address these issues, the study proposed two text prediction functions as data entry aids to system users. With 52 subjects, a two-group randomized experiment was conducted to quantify the impacts in terms of the reporting efficiency, quality, and usability attitudes. Consequentially, on structured data entry, the results were an overall 13.0% time reduction and 3.9% increase of response accuracy with the functions; on unstructured data entry, there was an overall 70.5% increase in the text generation rate, a 34.1% increase in the reporting completeness score, and a 14.5% reduction on the amount of text fields ignored by subjects. Subjects' usability attitudes were slightly improved with the proposed functions according to the questionnaire results.

Keywords:

Ppatient Safety Event Reporting; Usability; Text Prediction; Performance Measurement; Randomized Experiment.

Introduction and Background

In 1999, the prestigious report "To Err is Human" released by the Institute of Medicine estimated 44,000 – 98,000 patient deaths each year due to preventable medical errors [1]. In a recently published study in 2013, the estimation was raised to 210,000 – 440,000, which made the medical errors the third-leading cause of death, behind heart disease and cancer in the United States [2]. To learn from the errors in order to improve patient safety and healthcare quality, electronic event reporting (e-Reporting) systems have been proposed and promoted nationwide. However, low-quality reports pervade the systems and undermine the user engagement of e-Reporting.

As one of the most prominent factors associated with low-quality reports, data entry has received much attention in clinical information systems, such as electronic health record and computerized provider order entry system [3,4]. Quality data in patient safety holds promise in prompting system acceptance, which in turn may form a virtuous loop in leveraging system performance, user engagement, and patient safety. In this two-group randomized study, we explored text prediction techniques for facilitating data entry in terms of quality and efficiency.

Method and Material

Subjects

The study enrolled 52 nurses from 21 clinical departments in Tianjin First Central Hospital (TFCH) in China. All nurses were female between the ages of 30 and 52 years. On average, they had 20 years of nursing experience and four years of experience with e-Reporting. All participants signed an informed consent form approved by the Ethics Committee at the TFCH (No.E2012022K). This study was also approved by the Institutional Review Board at the University of Texas Health Science Center at Houston (No. HSC-SBMI-12-0767).

Interfaces

Two experimental interfaces (control and treatment) were developed as a means of configuration control during data collection. The two interfaces were identical in terms of task and representation, requiring the completion of 13 structured multiple-choice questions (MCQs) defined by AHRQ Common Formats [5] and one free text, multiple-line comment field for reporting additional details of patient fall. A single exception between the two interfaces was the provision of text prediction functions in the treatment interface, which included a cueing list (CL) and autosuggestion (AS). As illustrated in Figure 1, part B, the CL was attached to the MCQs (4 out of 13) that had a single text field. For unstructured data entry, the comment field was equipped with both the CL (part C) and AS (part G) in the treatment interface.

All items contained in the CL and AS were manually prepared [6]. Two domain experts reviewed all testing cases, and transcribed and categorized the key elements with the colloquial language in Chinese with characters Hanzi and phonetic typing method Pinyin. Pinyin text input methods (IME) are based on the standard qwert keyboard and are the most popular in Chinese. Pinyin IME does not require additional training for those who recognize English letters.

To prepare the CL and AS, expert reviewed words, phrases, and sentences were populated into a data table serving as reminder items. The number of listed items in any CL and AS was limited to 10, based on a trade-off between inspecting efforts and predicting sensitivity [7]. In the CL, at least one of the items was considered to be more descriptive, accurate, and pertinent to the case being reported than the rest. In the AS, the display of suggested entry options relied on a Soundex algorithm (phonetic matching function) of MySQL and the reporter's initial entries. As illustrated in Figure 1, Part G, the top 10 hits were rendered in the AS list. On the treatment interface, the subjects were allowed to select mixed entries from the AS list and user-generated text input. On the control interface, subjects were not provided any AS or CL; thus, each participant had to generate his or her own text input through a regular qwerty keyboard.

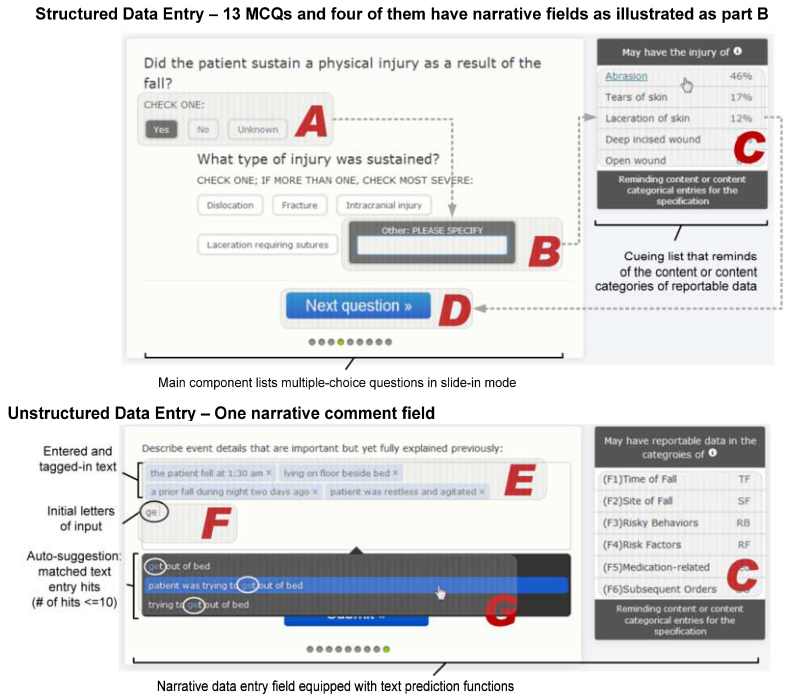


Figure 1 – The layout of interface elements is shown in English for both structured and unstructured data entries. Text prediction of *Cueing Lists* (CL, box C) and *Auto Suggestion* (AS, box G) are triggered with cursors in the corresponding fields. Box A shows the first answer triggering generated by the child's answers to questions B, C, D, etc. Box C is activated only when a single-text field (B) is open. CL reminds reporters of the content or key characteristics of reportable data associated with the event. The number of items listed in CL is <6. When pressing the button (D), a report would show an unstructured page containing boxes E, F and G. When initial letters (shown in box F) of the description are typed in, a matched list provided by AS shows up. The number of items listed in AS is <10. Matched letters and the focused line are highlighted in blue (G). Reporter may select one at a time and modify as necessary. Pressing the "Enter" button (behind G) will tag the current entry in a blue text chunk (as shown in E).

Testing Cases

In the study, each subject reported five patient-fall cases in a randomized sequence. The cases were selected from two sources —a case depository with 346 fall reports from a previous study [8] and a public database of Morbidity and Mortality (M&M) [9]. The five selected cases were translated into Chinese and rephrased by two domain experts for the purpose of quality and readability of text. A similar complexity of the five cases was agreed on by the domain experts. The following English narrative is an excerpt from one of the cases.

> *"… patient was alert and oriented X3 (person, time and location) upon assessment, and instructed on admit not to getting up without assist. He had been sleeping and attempted to get up to go to the bathroom. He forgot to call staff to have plexipulses (a device) undone, and tripped on plexi tubing and attempted to catch self on overhead bars. He landed on the floor…"*

Experimental Design

With a permuted-block algorithm, the 52 subjects were randomly assigned to two groups. 25 subjects were allocated into the group using the control interface without text prediction; 27 were assigned to the group with the treatment interface. The presenting sequence of five cases for each subject was randomly assigned. The subjects were trained through verbal instructions, and then asked to practice using both interfaces to report a sample case until each felt comfortable with the content and interface interactions. Since the training was conducted prior to grouping and the grouping procedure was blinded to both the subjects and the trainer, this design prevented confounding influences and training bias which might be delivered by the trainer.

Table 1 – The data sources and the units of analysis of key measures in the study

Measures	Data sources	Unit of analysis
Subject		
Age	Hospital nursing office	Years
Proficiency of reporting falls	Graded prior to the experiment	5 points Likert scale (1-low to 5-high)
Reporting efficiency		
Structured data entry	Accumulated time on MCQs	Seconds
Descriptive comments	Completion time on the comment field	Seconds
Text generation rate	Nominator: Letters in length of the comments; Denominator: completion time	Letters/Second*
Quality of reports		
Structured entry accuracy	Nominator: Accumulation of scores on MCQs; Denominator: Maximum of the accumulation	Percentage
Narrative completeness	The number of credited text chunks	Counts
Survey usability attitudes		
User attitudes in four dimensions	Posttest questionnaire	5 points Likert scale (1-low to 5-high)

* To count the length in letters, one UTF-8 encoded Chinese character equals three English letters in length.

In the hospital, a typical scene of event reporting is that a reporter initiates a report upon a witness's word-of-mouth information. This study simulated the natural scene by using the five cases with each having appeared on the first page of the interface. Subjects read the descriptions and answered the questions upon recall.

Pauses and pop-up questions were discouraged, except when a subject was in transition between reports. Keystroke level operations (mouse clicks and keystrokes) for each subject trial were time stamped and logged into a MySQL database. All reporting sessions were recorded using Camtasia Studio® 7 for data reconciliation. Each session was concluded with a questionnaire [10] via SurveyMonkey to reflect user attitudes in e-Reporting. The questionnaire was developed following Nielsen's Attitudes of Usability with a five-point Likert scale, wherein "1" indicated a maximal level of disagreement to the statement and "5" indicated a maximal level of agreement.

Processing of Data

The study collected ordinal and nominal data out of three data sources, including MCQs, narrative comment fields, and questionnaires. Ordinal data are the selected response items in the MCQs and questionnaires, and nominal data are the text entries in the single-line fields of MCQs and the comment field at the end of each report. We examined the ordinal and nominal data in terms of efficiency, effectiveness, and usability attitudes. Other experimental features associated with the CL and AS functions were also investigated. Table 1 lists the data sources and units of analysis of the measures.

Results

All the subjects successfully completed the experimental sessions with each subject having reported five cases, and the group consequentially generated 260 reports and 52 questionnaires. On average, each subject's session took 71 minutes, which consisted of 17 minutes in training and practice, 45 minutes in case-documenting, and 9 minutes in the questionnaire. There were 25 subjects in the control and 27 subjects in the treatment groups, accounting for 125 and 135 reports respectively. The mean ages of the subjects' groupings were 43.6 ± 5.8 (control) versus 41.1 ± 6.7 (treatment). The differences of ages and proficiency scores between the groups were insignificant ($p > 0.05$). The 260 reports contained 2,849 answers to MCQs and 238 unstructured narrative comments. As shown in Table 2, the subjects had eight significant variations between the groups indicated by an arrow ↑ (increase) or ↓ (decrease). Almost all variables, except for mouse clicks, in the treatment group showed an improved performance against the control group.

The analysis showed that the efficiency and quality of reporting can be enhanced by CL and AS for both structured and unstructured data entries. As illustrated in Figure 2, for the structured questions, the completion time was reduced by 13.0%, while the accuracy had a 3.8% increase overall. For unstructured narrative entries, the text generation rate was increased by 70.5%, which significantly increased the generation of free text data volume without an increase of time consumption between the groups. The reports generated in the treatment group were much more complete (34.2%) and richer in text (24.4%) than those in the control group.

As indicated in Table 2, the introduction of CL greatly improved the missing comments in the narrative field where the non-adherence rate dropped from 14.5% to 1.5%. CL functioned as a dynamic display when the page containing narrative comment fields slid in, as shown in Figure 1 (Part E, F & G). CL successfully drew subjects' conscious attention to the interface content. This dynamic CL signaled a compelling message to the subjects about the importance of filling the comment field.

Analytical results out of the questionnaires show potential non-adherence that may occur due to the following reasons: (1) a slip of unconscious skip of the comment field, (2) a lack of ideas of what event characteristics should be described further, and (3) that memory fades when working on the comment field.

Benefits from AS and CL also showed in keystrokes. In the reports, keystrokes were reduced by an average of 48.4%. In contrast, there was a 29.3% increase of mouse clicks. AS, in addition, showed a significant effect in generating 133 narrative comments, wherein AS was selected in 120 (90.2%) comments, totaling 460 times. Meanwhile, AS was inserted into 66.9% text chunks identified in the 133 entries. On average, AS was presented in 3.8 ± 1.9 text chunks in each report. A regression analysis showed that influential rate, defined as AS selection over total text chunks in each report, increased along the experimental progress ($p < 0.05$). Therefore, an observed learning effect amplified the efficiency along the experimental progress, as shown in Figure 4.

Fifty-two questionnaires were completed, which contained 1,300 rated answers. According to the analysis shown in Figure 5, the subjects showed overall good attitudes for the usability of both interfaces. The scores on all four dimensions slightly increased in the treatment group compared to the control group, but were not statistically significant. The text prediction functions implemented in the treatment group did not impact user attitudes in terms of learnability, efficiency, memory, and satisfaction.

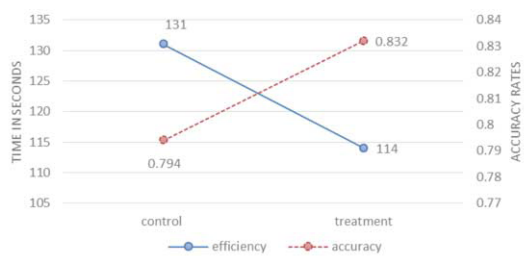


Figure 2 – Reporting efficiency and accuracy of structured data entries increased in the treatment group.

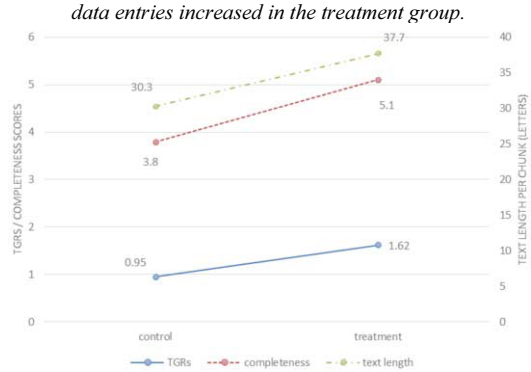


Figure 3 – Text generation rate and data completeness of unstructured data entries increased in the treatment group.

Table 2 – The subjects' performance in the reporting task.

Measure	Control Group	Treatment Group	Variation	*p*-value	Illustration
Subject					
Sample size (subjects/reports)	25/125	27/135			
Age	43.6±5.8	41.1±6.7		0.189	
Proficiency in reporting fall events (Likert scale 1-5)	3.8±0.5	3.9±0.5		0.413	
Reporting efficiency					
Time on case reading (seconds)	84.9±44.8	78.6±43.2		0.307	
Structured data entry (seconds)	131.0±50.0	114.0±41.7	↓ 13.0%	0.004	Figure 2
Descriptive comments (letters)	139.6±99.6	142.9±82.2		0.782	
Text generation rate (letters per second)*	0.95±0.35	1.62±0.99	↑ 70.5%	0.000	Figure 3, 4
Physical operations in a report					
Mouse click	13.3±2.1	17.2±3.7	↑ 29.3%	0.000	
Keystroke	173.2±117.0	89.4±89.4	↓ 48.4%	0.000	
Quality of reports					
Structured entry accuracy	79.4±10.1%	83.2±11.0%	↑ 3.8%	0.000	Figure 2
Missing comments (non-adherence rate)	20/125(16.0%)	2/135(1.5%)	↓ 14.5%	0.000	
Narrative completeness	3.8±2.3	5.1±2.4	↑ 34.2%	0.000	Figure 3
Length of a chunk - richness (letters)*	30.3±13.1	37.7±18.6	↑ 24.4%	0.000	Figure 3
Survey usability attitude (Likert scale 1-5)					
Learnability	4.0±0.3	4.1±0.5		0.545	Figure 5
Efficiency	4.1±0.4	4.3±0.6		0.386	Figure 5
Memory	3.5±0.5	3.8±0.5		0.099	Figure 5
Satisfaction	3.8±0.4	3.9±0.5		0.458	Figure 5

* To count the length in letters, one UTF-8 encoded Chinese character equals three English letters in length.

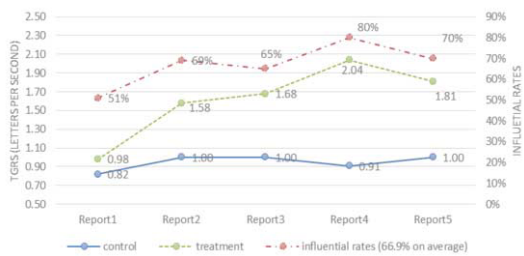


Figure 4. Text generation rates exhibit an overall increase (learning effect) over five reports in the treatment group. In contrast, such an effect does not present in the control group.

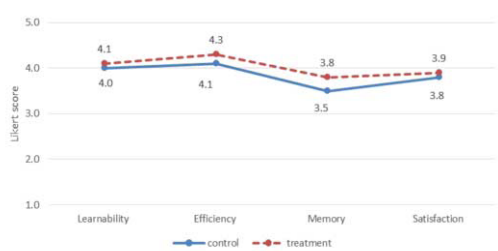


Figure 5. User attitudes of the treatment group are slightly improved than the control group.

Discussion

It is generally agreed that keystrokes could be reduced when clickable on-screen options are readily available on a user-centered interface. However, the keystroke savings alone remain inadequate in explaining the increased efficiency. Earlier research exhibits mixed, inconsistent results regarding improved efficiency of keystrokes in terms of cognitive load, eye gaze movement, and mouse clicking [11,12]. The balance between mouse clicks and keystrokes is an interesting topic in the process of developing an efficient and effective e-Reporting system. As touchscreen handheld devices are becoming pervasive in healthcare, significant keystroke savings on data entry may greatly amplify the benefits of user-centered design. This is because most touchscreen users prefer on-screen menu selection to on-screen keyboard typing.

The two-group randomized experiment proved the effectiveness of AS and CL text prediction in generating quick and high quality data entry in e-Reporting. User attitudes reflected in the questionnaires did not change significantly, but were potentially improved with the introduction of text predictions in e-Reporting. The observed effect may be generalizeable to other data entry interfaces in healthcare. Clinicians working under time pressure are expected to document in a timely manner [13,14]. Meanwhile, the quality of data has to achieve a high standard for decision-making and the creation of actionable knowledge. The solutions such as AS and CL experimented with in this study pave a way for addressing these concerns. The clinical document architecture (CDA) is a markup standard developed by Health Level 7 Internation (HL7) to define the structure of clinical documents, such as discharge summaries, progress notes, etc. Nonetheless, e-Reporting is becoming a routine documentation task where the structure is not yet well defined.

Recently, The Agency for Healthcare Research & Quality (AHRQ) created the Common Formats (CF), which are common definitions and reporting formats, to help providers uniformly report patient safety events and to improve healthcare quality. CFs include paper forms to guide the development of data collection intruments. When transforming the paper forms into electronic interfaces, designers should carefully learn from those lessons based upon the evolvement of electronic medical records.

One of the major purposes of e-Reporting is the collection of information and knowledge about patient safety events. Ironically, the e-Reporting process is not immune to human errors due to time pressure, multi-tasking, or competing priorities.

Recommendations for Interface Development in Healthcare

Guided by user-center design principles, we employed state-of-the-art and open-source techniques to rapidly prototype the dual-interfaces, one exhibited stereotypical functions in prevailing e-Reporting, the other was renovated and embeded with text predictions. The dual interfaces not only met the experimental requirements of randomization and control, but successfully demonstrated the effectiveness and effiency of the innovative design in e-Reporting. Interface design specifically for the purpose of e-Reporting requires further investigation, which would contribute to the purpose of collecting quality events, generating actionable knowledge, and sharing experience in reducing the recurrence of similar incidents.

Our design and development process is instrumental to many clinical information systems. The incremental effect on saving providers' time and improving report quality may be magnified, while evident-based estimates show 440,000 preventable events each year in the U.S. [2].

Limitations

CL offered in the interface presents a high quality contribution by domain experts. In reality, the prediction accuracy based upon an algorithm of event similarity and frequency of appearance may not match the manual effort of experts. In addition, the number of predicted items may differ in other clinical systems depending upon the richness of data entries. We did not provide an extended item list, which may be a burden for reporters' inspection. In the future, the question of whether text prediction with lower accuracy or a longer list would undermine the proven effect is worth further investigation.

It is a noteworthy dilemma that AS, designed for increasing text generation rate and accuracy, may lead to side effects, such as a subject's overdependence on the suggested items, and thus limit constructive thinking [15-17]. This is because, when provided with a predefined list of options, busy clinicians tend to recognize or ignore the options rather than specifying answers. The future development of the AS should carefully keep a balance between the strength and weakness of such protocols, and strategically address the dilemma.

Conclusion

User-centered design aimed at renovating e-Reporting demonstrated the effectiveness of text prediction in clinical data entry. This work may lay the groundwork for further research associated with text prediction in clinical information systems.

Acknowledgments

The authors express heartfelt thanks to Shen Wang, RN, Xiao Liu, RN and Ying Sun, RN at the Office of Nursing at Tianjin First Central Hospital in China who helped review and edit the testing cases, derived proper text prediction responses, scheduled the testing sessions, and graded the subjects' commentary data.

References

1. Kohn LT, Corrigan JM, Donaldson MS. To err is human: building a safer health system. Report of Committee on Quality of Healthcare in America. Institute of Medicine, National Academy of Science, 1999.
2. James JT. A new, evidence-based estimate of patient harms associated with hospital care. Journal of Patient Safety 2013: 9 (3): 122-8.
3. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. Journal of the American Medical Informatics Association 1997: 4 (3): 213-21.
4. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ 2004: 328 (7449): 1184-7.
5. AHRQ. Common Formats 2011 [Internet]. [Place unknown]: AHRQ; 2011 [cited _____give date YYYY MMM DD_____]. Available from: http://www.psoppc.org/web/patientsafety/version-1.2_documents
6. Higginbotham DJ, Bisantz AM, Sunm M, Adams K, Yik F. The effect of context priming and task type on augmentative communication performance. Augmentative and Alternative Communication 2009: 25 (1): 19-31.
7. Hunnicutt S, Carlberger J. Improving word prediction using markov models and heuristic methods. Augmentative and Alternative Communication 2001: 17 (4): 255-64.
8. Gong Y. Terminology in a voluntary medical incident reporting system: a human-centered perspective. Proceedings of the 1st ACM International Health Informatics Symposium (1882996) ACM, 2010: 2-7.
9. AHRQ. Web M&M mobidity and mortality rounds on the web:cases and commentaries [Internet]. [Place unknown]: AHRQ; 2011 [cited 2011 Dec 4]. Available from: http://www.webmm.ahrq.gov/home.aspx
10. Hua L. Posttest questionnaire for user reflections on the system 2013 [Internet]. [Place unknown]: Name Site; date [cited___YYYY_MMM_DD___] Available from: http://www.surveymonkey.com/s/9SLCZ9M
11. Light J, Lindsay P, Siegel L, Parnes P. The effects of message encoding techniques on recall by literate adults using AAC systems. Augmentative and Alternative Communication 1990: 6(3): 184-201.
12. Koester HH, Levine S. Effect of a word prediction feature on user performance. Augmentative and Alternative Communication 1996: 12(3): 155-68.
13. Allan J, Englebright J. Patient-centered documentation: an effective and efficient use of clinical information systems. Journal of Nursing Administration 2000: 30(2): 90-5.
14. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. Journal of the American Medical Informatics Association 2005: 12(5): 505-16.
15. Gong Y. Data consistency in a voluntary medical incident reporting system. Journal of Medical Systems 2009: 35(4): 609-15.
16. Tamuz M, Harrison MI. Improving patient safety in hospitals: contributions of high□reliability theory and normal accident theory. Health Services Research 2006: 41(4p2): 1654-76.
17. Tamuz M, Thomas E, Franchois K. Defining and classifying medical error: lessons for patient safety reporting systems. Quality and Safety in Health Care 2004: 13(1): 13-20.

Address for Correspondence

Dr. Yang Gong. 7000 Fannin St, Houston, Texas, USA 77030.
yang.gong@uth.tmc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-193

The Role of Hospital Information Systems in Universal Health Coverage Monitoring in Rwanda

Gustave Karara[a], Frank Verbeke[a,b], Marc Nyssen[a,b]

[a] *Department of Biostatistics and Medical Informatics,VrijeUniversiteit Brussel, Brussels, Belgium*
[b] *Regional e-Health Center of Excellence, University of Rwanda, Kigali, Rwanda*

Abstract

In this retrospective study, the authors monitored the patient health coverage in 6 Rwandan hospitals in the period between 2011 and 2014. Among the 6 hospitals, 2 are third level hospitals, 2 district hospitals and 2 private hospitals. Patient insurance and financial data were extracted and analyzed from OpenClinic GA, an open source hospital information system (HIS) used in those 6 hospitals. The percentage of patients who had no health insurer globally decreased from 35% in 2011 to 15% in 2014. The rate of health insurance coverage in hospitals varied between 75% in private hospitals and 84% in public hospitals. The amounts paid by the patients for health services decreased in private hospitals to 25% of the total costs in 2014 (-7.4%) and vary between 14% and 19% in public hospitals. Although the number of insured patients has increased and the patient share decreased over the four years of study, the patients' out-of-pocket payments increased especially for in-patients. This study emphasizes the value of integrated hospital information systems for this kind of health economics research in developing countries.

Keywords:

Hospital information system; universal health coverage; monitoring health insurance; patient out-of-pocket payments; Rwanda.

Introduction

Providing universal access to health care remains a challenge for many low income countries especially because of the weakness of their health financing systems [1,6]. In the Sub-Saharan African context, Rwanda currently sustains one of the most elaborate health insurance schemes. Health care services are funded by both public and private health insurance organizations. Public health insurance schemes include the Rwanda Social Security Board (RSSB) and the Military Medical Insurance (MMI) which cover government employees and the Community Based Health Insurance (CBHI) schemes, also called Mutuelles de Santé, which cover the majority of the remaining population (according to the Ministry of Health, in 2012, more than 90% of the population was covered by CBHI [2]). Additionally, private health insurance (PHI) schemes are offered by a growing number of private health insurers. Public and private health insurance organizations have different objectives and therefore provide different levels of health insurance coverage to different groups of affiliates. CBHI schemes cover the majority of the population in the informal sector and those who are categorized as poor, very poor or indigents. RSSB and MMI schemes cover, respectively, the staff of the Public Service and staff of Security Services (Army and Police) and their beneficiaries. PHI schemes cover patients from the formal sector and liberal professionals according to the level of insurance of their choice [3].

In general, CBHI schemes cover 90% of provided care delivery costs. RSSB and MMI cover 85% and PHI coverage varies between 85% and 100%. CBHI only covers health services provided by the public health facilities while affiliates of RSSB, MMI and PHI also have access to private health care.

Rwandan health insurance companies can freely negotiate tariffs with public and private health facilities for the care deliveries they wish to cover but the Ministry of Health can always enforce fixed tariffs for essential health services. Four tariff categories are used in Rwanda: (1) CBHI, (2) RSSB/MMI, (3) PHI and (4) people without health insurance. Invoice processing, payment status tracking and health care reimbursement rules have become increasingly complex [4]; some health facilities have therefore chosen to implement a Hospital Information System (HIS) to manage these activities.

From 2007 until 2014, the authors participated in the implementation of the open source OpenClinic GA HIS [4,5] in 17 public and private clinics and hospitals in Rwanda. OpenClinic GA is an integrated hospital information management system resulting from a research project at the Vrije Universiteit Brussel (VUB) that includes most of the modules commonly found in modern HIS [5], such as:

- Patient administration
- Admission, discharge and transfer (ADT) management
- Financial information management
- Reason for encounter and diagnostic coding
- Medical record management
- Lab information management
- Medical imaging
- Reporting and statistics

The software was developed in Java connecting over JDBC to the most popular ANSI SQL 92 compliant database servers (such as MySQL and MS SQL Server) and offers an easy to use web interface facilitating HIS deployment in often challenging technological settings commonly found in developing countries [4,5]. OpenClinic GA has been built according to the GEHR (Good Electronic Health Record) architecture [4]. The data model is patient oriented.

Currently OpenClinic GA is being used in more than 500 health facilities in 84 countries and 45 of these sites in Rwanda, Burundi, DR Congo, Mali, Senegal, Gabon and Congo Brazzaville are being closely monitored by the VUB ICT4Development research group [12]. OpenClinic GA has

gradually become the first medical Open Source software project downloaded over the world [11].

Using patient insurance and financial data available in Openclinic GA systems from 6 Rwandan hospitals, and in order to determine the level of population health coverage in the hospital, this study aims to evaluate: (1) the percentage of patients with health coverage, consulting health facilities in Rwanda, (2) to what extent patients make use of the possibility to get private healthcare when this is covered by their health insurance plan, (3) the evolution of the level of reimbursement of different health service categories and (4) the evolution of patient payments for uncovered health services. This descriptive and analytical retrospective study was performed in 2 third level referral hospitals, 2 district hospitals and 2 private hospitals.

Methods

The study was conducted in 6 Rwandan hospitals:

- 2 national referral hospitals: the University Teaching Hospital of Kigali (CHUK) and the Neuro-Psychiatric referral Hospital Caraës-Ndera (Kigali) (NPH-CN).
- 2 district hospitals in the Eastern Province in the cities of Nyamata (DHNY) and Rwamagana (DHRW)
- 2 private hospitals in Kigali: La Croix du Sud Hospital (CDS) and La Médicale Polyclinic (LMED).

OpenClinic GA has been installed and implemented in these 6 hospitals since 2010 and was configured to cover all possible statuses of patients' health insurance. The financial module of the system manages the various patient health insurance schemes, care deliveries consumed by patients and all billing operations for patients and their insurers. We analyzed the health coverage information available in the HIS health insurance module for all patients treated between 1/1/2011 and 30/11/2014.

First, we isolated two patient groups from the database: those who presented to the hospital with at least one valid health insurance and those who had no health coverage. We then evaluated the evolution of health insurance coverage during the study period.

Secondly, for each outpatient visit and each inpatient admission, we determined the type of health coverage that applied. Four types of health service payments were distinguished: (1) patient direct payments (PATIENT), (2) public health insurance payments (RSSB/MMI), (3) private health insurance payments (PHI) and (4) community health insurance payments (CBHI). We then calculated the weighted average percentages for the different types of health coverage in those hospitals and compared them using the Chi² test in the Center for Disease Control's Epi Info.

Finally, we extracted the amounts paid by patients and insurers in each hospital and assessed the evolution of their proportions and the absolute amounts paid by the patients for health services (patient out-of-pocket payments).

Results

Health services coverage

We analyzed a total of 778,915 electronic records of patients who consulted the 6 hospitals. The distribution of patient numbers is shown in Table 1.

Table 1 - Number of Electronic Patient Records Analyzed

CHUK	NPH-CN	DHRW	DHNY	CDS	LMED	Total
216 389	32 128	92 907	89 218	170 852	177 421	**778 915**

The evolution of the proportion of patients who are covered by at least one health insurance scheme shows a steady increase in all six hospitals in the course of the 4 year study period, as shown in figure 1. In the Neuro-Psychiatric hospital of Ndera (NPH-CN), the health insurance coverage increased from 90% to 98% between 2011 and 2014 while at CHUK, it grew from 74% to 85%. In the district hospitals of Nyamata (DHNY) and Rwamagana (DHRW), health insurance coverage improved from 78% to 88%. Similar results were obtained from private hospitals, where health coverage grew from 63% to 72% at La Médicale (LMED) and from 70% to 76% at La Croix du Sud (CDS).

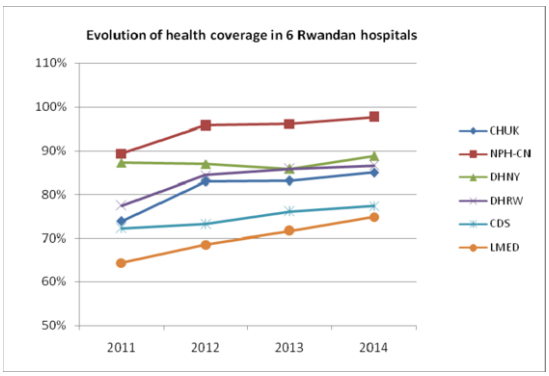


Figure 1 - Evolution of patient health coverage in 6 hospitals

Health insurance schemes in hospitals

For every out-patient encounter or in-patient admission, part of the health service costs must be paid by the insurer and the remainder by the patient, according to the patient's health coverage plan. A number of health services remain uncovered by some health coverage plans and therefore patients' out of pocket payments may vary considerably depending on the type of care that was provided. In this step, we analyzed the real reimbursement levels of different health insurance schemes over time, taking into account the uncovered parts of care.

A total of almost 2.9 million encounters (2,534,967 outpatients and 342,878 inpatients) have been analyzed, as shown in Table 2.

Table 2 - Numbers of Encounters Analyzed

	In-patient	Out-patient	Total
CHUK	92 103	639 447	731 550
NPH-CN	22 260	244 653	266 913
DHRW	110 941	224 817	335 758
DHNY	58 665	260 176	318 841
CDS	55 655	708 327	763 982
LMED	3 254	457 547	460 801
Total	342 878	2 534 967	2 877 845

The number of patients who were hospitalized at LMED is low because the use of OpenClinic GA for hospitalizations is very recent in this hospital. CHUK and CDS are 2 reference hospitals in Kigali, respectively, for the public and the private health sector. They contribute to a significant part of in- and

out-patients in our study. Figure 2 and figure 3 show the average proportion of insurance schemes coverage, respectively, for in-patients and out-patients in the 6 hospitals during the study period.

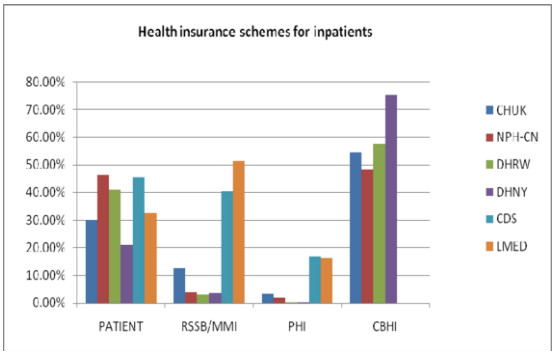


Figure 2 - Health insurance schemes for in-patients

In general, for in-patient encounters, the community based health insurance (CBHI) is the most commonly used scheme (weighted average: 60.7%) followed by public insurance (RSSB/MMI) with a weighted average of 27.7%, private insurance (PIH) for 12.9% and finally the patient directly paying for uncovered health services (PATIENT) in 37.6% of the cases. In over 45% of the in- patient encounters, patients paid directly for health care at NPH-CN and CDS due to certain care deliveries (especially hygiene supplies, specialty medicines) not being covered by the insurer. We noticed that the patient used direct payments for uncovered health services more frequently ($p<0.0001$) in the in-patient encounters in the private hospitals (44.9%) than in district hospitals (37.9%) and tertiary public hospitals (34.5%). The public health insurance coverage scheme (RSSB/MMI) is most commonly used for in-patient encounters in private hospitals (51% at LMED and 40% at CDS). There is a statistically significant difference ($p<0.0001$) between using public health insurance (RSSB/MMI) in private hospitals' (41.20%) and in public hospitals' (9.43%) in-patient encounters. At CHUK the RSSB/MMI scheme is used in 12.7% of in-patient encounters. The CBHI scheme is only used in public hospitals. It is used more frequently ($p<0.0001$) in district hospitals (64.85%) than in third level hospitals (53.38%). Finally, patients with private health insurance (PHI) are more often hospitalized in private hospitals (16.8% of the in-patients) than in the public ones (2.8% of the in-patients).

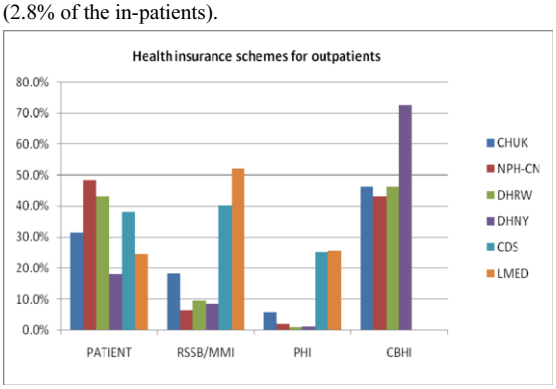


Figure 3 - Health insurance schemes for out-patients

Out-patient health coverage schemes follow a similar distribution. The CBHI scheme is more used (53.0% of the encounters) in out-patient encounters than other health insurance schemes: RSSB/MMI (38.0%), PHI (22.4%) and PATIENT (24.6%). Out-patients are facing more uncovered health services in public tertiary hospitals (37.8%) than in district hospitals (35.2%) and private hospitals (34.2%). The RSSB /MMI scheme is more used in private (45.8%) than in public health facilities (14.9%). The CBHI scheme is also more used in district hospitals (63.2%) than in tertiary hospitals (45.5%). Patients with private insurance were more likely to consult private (25.3%) than public hospitals (5.0%).

Multiple insurance schemes can co-occur in the same encounter for the same patient.

Health services payment in hospitals

Between 2011 and 2014, the proportion of patient out-of-pocket amounts paid, compared to the total amounts paid for the health service costs, has declined significantly in private hospitals: from 36.8% to 27.6% at LMED (-9.3%) and from 29.9% to 24.4% in CDS (-5.5%). This proportion also decreased at CHUK from 21.6% to 17.5% (-4.1%) but remained stable at NPH-CN (14.2%) and increased slightly in district hospitals: from 13.2% to 15.0% at DHNY (+ 1.8%) and from 17.2% to 19.7% at DHRW (+ 2.5%). Figure 4 shows the evolution of patient out-of-pocket payment proportions for health services between 2011 and 2014 in the 6 hospitals.

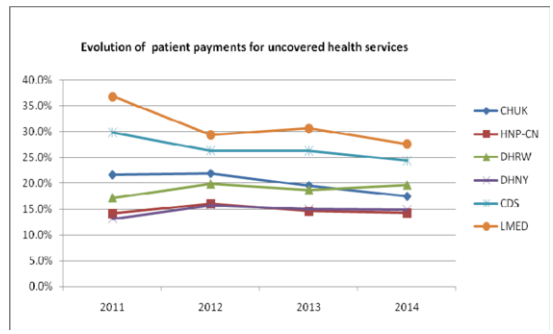


Figure 4 - Evolution of patient payment proportions for health services

The reason for the patient out-of-pocket payment proportion increase at DHRW is found in the increase of some non-reimbursable consumables (+ 2.6%), acts (+ 2.8%) and hospital stay (+ 2.6%).

Although the proportion of the amounts paid by patients has generally decreased during the study period, the absolute patient out-of-pocket payments for health services show a different trend as is demonstrated in figures 5 and 6.

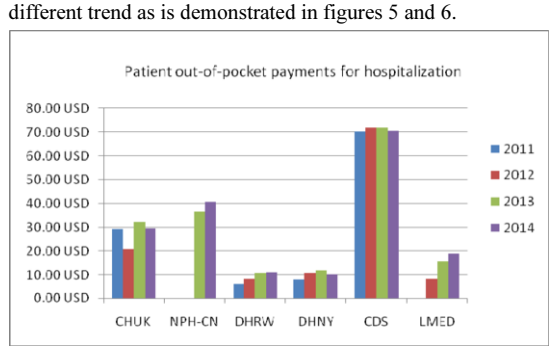


Figure 5 - Average amounts paid by the patient for health services for hospitalization

For hospitalization, the patient out-of-pocket amounts disbursed for health services at CHUK decreased the first year from 29.27USD to 20.67USD but then increased again. In 2014, the patient paid an average of 30.00USD for his hospital stay. At NPH-CN, patients' average payments for the hospitalization were 40.60USD in 2014 (NPH-CN began to use the invoicing module better in 2013). In district hospitals, patient out-of-pocket payments increased from 5.95USD at DHRW and 7.90USD at DHNY to 11.00USD and 10.15USD respectively. We noticed a large difference between patient payments in the two private hospitals: At CDS, the patient was paying around 70USD per hospitalization, while at LMED the patient out-of-pocket payments increased from 8.34USD to 18.88USD (LMED started to use the invoicing module well in 2012). The CDS experienced a period during which health care reimbursements were refused by some insurers before their accreditation as a "hospital" because their health service tariffs (especially for the hospital rooms) were too high (and the majority of the costs were not reimbursed by insurers) compared to other private hospitals and clinics.

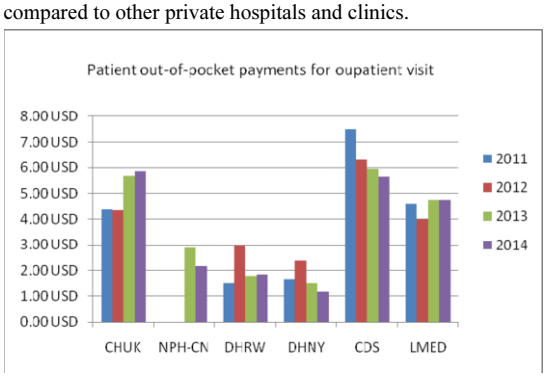


Figure 6 - Average amounts paid by the patient for health services in out-patient encounters

For out-patients between 2011 and 2014, patient out-of-pocket payments tended to increase at CHUK (from 4.38USD to 5.87USD), remain stable at LMED (around 4.50USD), and to decrease from 2012 in district hospitals (from 2.93USD to 1.83USD at DHRW and from 2.37USD to 1.15USD at DHNY), at NPH-CN (from 2.88USD to 2.19USD) and at CDS (from 7.49USD to 5.65USD). The decrease of patient out-of-pocket payments observed from the second year (2012) in district hospitals was explained by the increased number of insured patients and some consumables (gloves, syringes and needles) used during medical examinations which were initially charged to the patient and included in the global cost of the exams afterward. The decrease of patient payments observed at CDS was due to the increase of insured patients.

Discussion

This study focused on the monitoring of health coverage in Rwandan hospitals. The results showed that the health coverage in the 6 analyzed hospitals gradually increased between 2011 and 2014, reaching 98% at NPH-CN, 85% at CHUK, 88% in districts hospitals (DHRW and DHNY) and around 74% in private hospitals (LMED and CDS). Health coverage projections done in 2010 showed that in 2014, at least 92% of the population would have a valid health insurance for medical care in public health facilities [2]. This goal was reached at NPH-CN, a specialized national reference center for mental health care. Almost all patients treated at NPH-CN have a valid health insurance because the reference policy from district hospitals to tertiary hospitals enforces that insurance. In addition, difficult social cases are directly covered by the Ministry of Health. The Rwandan government pays special attention to psychiatric disorders which are often associated with higher financial risks [7,9] in a country that still faces psychological trauma consequences of 1994's genocide and war.

The high level of insured patients seen in private health facilities is explained by private health care coverage offered by public insurance schemes (RSSB / MMI) and private insurances (PHI) as shown in figures 2 and 3. Patients covered by public or private insurance which reimburse private health care are more likely to consult private hospitals where they hope to receive better quality of care [8,9].

The proportion of patients' out-of-pocket payments compared to total health service costs has declined in private hospitals and at CHUK. It remained almost stable in other hospitals (figure 4). But the average amounts actually paid by patients for health services tended to increase for in-patients and to decrease for out-patients. The decrease of patients' out-of-pocket payments is explained by the fact that the number of patients with valid health insurance has increased and insurers accept reimbursing certain health services that previously were not covered. This assertion was also confirmed by the study of Ties Boerma and colleagues [10]. However, this assertion is not confirmed for in-patients where certain health services remain to be reimbursed only partially or not at all by health insurers.

Conclusion

The use of an open source hospital information management system enabled the detailed monitoring of universal health coverage in 6 Rwandan hospitals. The results show improvement in health coverage in both public and private hospitals in the period between 2011 and 2014 due to the role of public, community and private health insurance schemes. But efforts are still needed to further reduce the patients' out-of-pocket payments.

This study demonstrates the possibility to assess the level of universal health coverage in a developing country using hospital management systems. Although we targeted referral health facilities, this experience could also be applied to first-level health facilities such as health centers through an adequate ICT infrastructure based health information management.

References

[1] WHO. World Health Statistics. 2014

[2] Minsitry of Health.. Rwanda Community based health Insurance policy. Kigali; April 2010. [accessed 8 april 2015]. Available from: http://www.moh.gov.rw/fileadmin/templates/Docs/Mutual_policy_document_final1.pdf

[3] Musango L, Doetinchem O, Carrin G. De la mutualisation du risque maladie à l'assurance maladie universelle Expérience du Rwanda. OMS 2009. [accessed 8 april 2015]. Available from: http://www.who.int/health_financing/documents/dp_f_09_01-mutualisation_rwa.pdf

[4] Verbeke F. ICT based health information management methods for supporting and improving health care in sub-Saharan hospitals; A field study in Rwanda, Burundi, Mali, Ivory cost and the Democratic Republic of the Congo. Thesis, VUB, 2012.

[5] Verbeke F. OpenClinic GA. [accessed 8 April 2015]. Available from: http://sourceforge.net/projects/open-clinic

[6] World Health Organisation. Health system financing: the path toward universal coverage. Geneva, 2010.

[7] Carrin, G, Mathauer, I, Evans, D (2008). Universal coverage of health services: tailoring its implementation. Bull World Health Organisation, 86; 857-863.

[8] World Health Organization. The World Health Report 2013. Research for Universal Health Coverage. [accessed 8 April 2015]. Available from: www.who.int/universal_health_coverage/en/

[9] Basumail S, Andrews J, Kishore S, Panjabi R, Stuckler D. Comparative performance of private and public healthcare systems in low- and middle-income countries: a systematic review. 2012 June 19; DOI: 10.1371/journal.pmed.1001244.

[10] Boerma T, Eozenou P, Evans D, Evans T, Kieny M, Wagstaff A. Monitoring progress towards universal health coverage at country and global levels. 2014 Sept 22; DOI: 10.1371/journal.pmed.1001731.

[11] Medical Free/Libre and Open Source Software. [accessed 9 April 2015]. Available from: http://www.medfloss.org/node/271

[12] Globalhealthbarometer. [accessed 9 April 2015]. Available from: http://www.globalhealthbarometer.net/globalhealthbarometer/datacenter/datacenterHomePublic.jsp

Address for correspondence

Gustave Karara, MPH, PhD Candidate

Department of Biostatistics and Medical Informatics

Vrije Universiteit Brussel

Laarbeeklaan 103

B-1080 Brussels

Belgium

gustave.karara@vub.ac.be

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-198

Bluetooth Roaming for Sensor Network System in Clinical Environment

Tomohiro Kuroda [a], Haruo Noma [b], Kazuhiko Takase [c], Shigeto Sasaki [d], Tadamasa Takemura [e]

[a] *Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan.*
[b] *Department of Media Technology, Ritsumeikan University, Kusatsu, Japan.*
[c] *Takebishi Corporation, Kyoto, Japan.*
[d] *Shimadzu System Development Corporation, Kyoto, Japan.*
[e] *Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan.*

Abstract

A sensor network is key infrastructure for advancing a hospital information system (HIS). The authors proposed a method to provide roaming functionality for Bluetooth to realize a Bluetooth-based sensor network, which is suitable to connect clinical devices. The proposed method makes the average response time of a Bluetooth connection less than one second by making the master device repeat the inquiry process endlessly and modifies parameters of the inquiry process. The authors applied the developed sensor network for daily clinical activities in an university hospital, and confirmed the stabilitya and effectiveness of the sensor network. As Bluetooth becomes a quite common wireless interface for medical devices, the proposed protocol that realizes Bluetooth-based sensor network enables HIS to equip various clinical devices and, consequently, lets information and communication technologies advance clinical services.

Keywords:

Sensor Network; Bluetooth; Roaming; Hospital Information System; Clinical Application;

Introduction

A sensor network is key infrastructure for advancing a hospital information system (HIS). A sensor network enables one to plug various devices including vital sensors into a HIS and lets the devices send obtained data directly to connecting HISs. Consequently, the sensor network frees clinical staff from posting vital data into a HIS and frees patients from clinical incidents caused by misposting of the data.

The sensor network enables us to divide complicated medical equipment, such as the echogram, into functional units and embed the equipment into a gigantic computer system, which is called hospital [1]. Embedded medical equipment with tiny sensing units, such as wireless ultrsound probes, frees clinical staff from carrying heavy medical equipment or patients to move to rooms with medical equipment, and enables the clinical staff to perform measurements anywhere. As a matter of course, obtained data can be directly stored to a HIS, and the data may enable a HIS to provide certain clinical diagnostic support in proper moment. Thus, embedding medical devices into a HIS can be a silver bullet to increase clinical productivity and safety by maximizing the benefit of information and communication technologies [2].

A sensor network can be realized by any wireless communication technology, such as WiFi, Zigbee, and Bluetooth. A recent increase of social consciousness of health promotes vital sign sensors to equip connectivity with mobile information terminals, such as smartphones. The most common standard to mediate communications among personal health devices is Continua Health Alliance [3], which defines the standard connection protocol via Bluetooth. Recently, the existence of the Continua promotes clinical sensors to equip Bluetooth. Thus, once Bluetooth equips roaming functionality for enabling sensors to move around and real time location functionality to locate those moving sensors, it can be a feasible technology for emerging clinical sensor network environment.

On the other hand, the Bluetooth protocol, which is originally designed as “wireless serial bus”, is not suitable for roaming. The conventional Bluetooth protocol requires a time-consuming “pairing” process to connect a host computational device with a client terminals such as sensors. In this paper, the authors propose a method to provide roaming functionality for Bluetooth, implement the proposed method as a commercial Bluetooth sensor network system, and evaluated the developed system during daily clinical tasks at the Kyoto University Hospital.

Methods and Materials

Figure 1 shows the state transition diagram of Bluetooth. A Bluetooth module goes into “Standby” phase when it is switched on. When a CPU sends an “Inquiry” command to the Bluetooth module in standby phase, the module moves into the inquiry phase and starts to search nearby Bluetooth devices for a master communication device. Once a Bluetooth slave device such as a headset or a keyboard catches an inquiry packet from the master communication device, it responds with a Frequency Hop Synchronization (FHS) packet. After exchanging the inquiry and the FHS packet, both master and slave devices move into the “Page” phase and exchange communication parameters. Finally, both master and slave device move into “Connect” and “Transmit” phases to start data exchange.

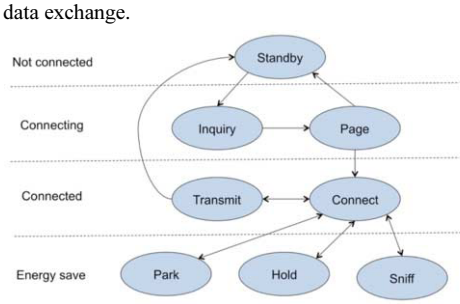


Figure 1 – The state transition table of Bluetooth

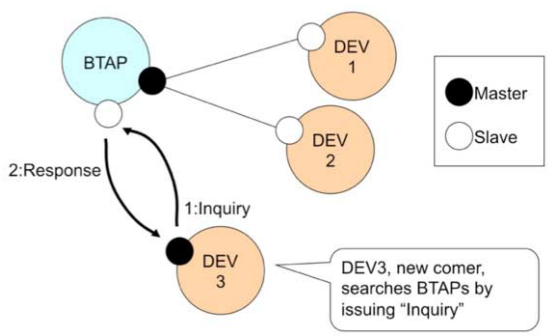
As a Bluetooth device moves into the page status immediately after discovery to wait for communication, a moving device, which is once coupled with a certain Bluetoth access point (BTAP), needs to wait for timeout to know that it is already out of reach of the BTAP. After that, the device needs to go through the inquiry process to find nearest BTAP. Moreover, the inquiry process is quite time consuming. The process requires approximately ten seconds. This limitation disables Bluetooth devices for smooth roaming under multiple BTAPs.

The authors made the master device suspend and restart the inquiry process in shorter periods by changing parameters, such as the master's inquiry interval and the slave's inquiry scan window size.When a master device suspends and restarts inquiry process every 640 milliseconds, the average response time to discover slaves becomes less than one second. As medical devices under operation in clinical environment will not move quicker than walking speed, the achieved response time, less than one second, is adequate to let devices move around.

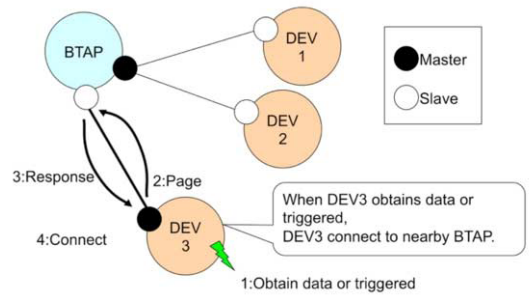
Once a device obtains certain data to send, the device should be paired with a BTAP within a reach. As the paring process initiated by the medical device, the proposed system makes the medical devices act as the master nodes during inquiry mode. On the other hand, to enable a BTAP to accept connections from multiple devices, the BTAP needs to act as the master node of the particular connection to keep slave channel open to accept incomming connection request.

Figure 2 shows the detailed protocol of the pairing process. In this diagram, a new comer (DEV3) tries to connect to a BTAP, which is connected with two devices (DEV1 and DEV2). In the beginning, DEV3 works as a master Bluetooth device to repeat the inquiry process as Figure 2(a) to keep finding nearest BTAP. When DEV3 obtains certain data to send, DEV3 immediately connects with the BTAP through the standard Bluetooth protocol as shown in Figure 2(b), then swaps their roles with the BTAP as shown in Figure 2(c). As a result, the BTAP keeps its slave channel open to accept an incoming device as shown in Figure 2(d).

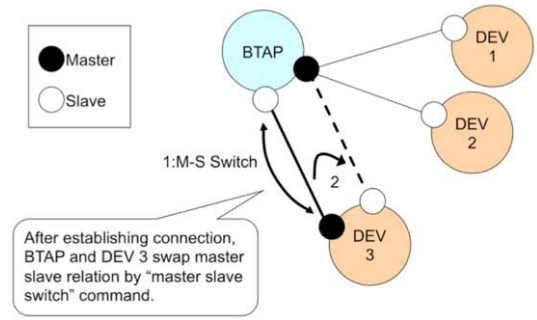
The BTAP keeps connection with DEV3 until DEV3 sends a termination code, an expected (pre-defined) data communication is finished, or the the BTAP detects timeout (Figure 2(e)). When DEV3 fails finishing communication, DEV3 goes back to inquiry process to find the nearest BTAP to restart the communication.



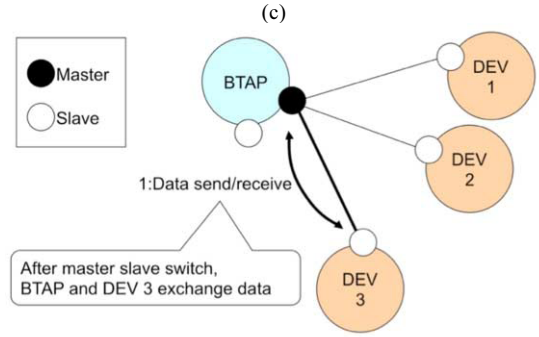
(a)



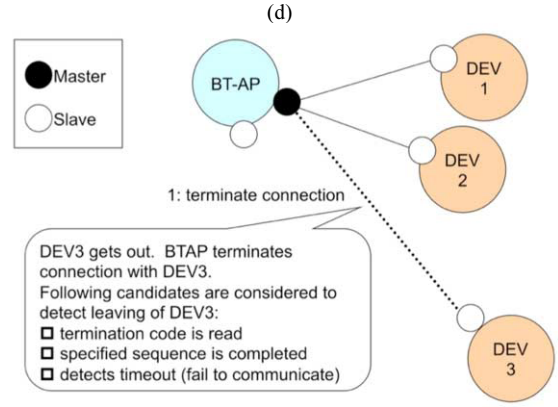
(b)



(c)



(d)



(e)

Figure 2 – The proposed pairing protocol (a) DEV3 keep finding BTAP by repeating inquiry phase. (b) DEV3 connects with BTAP. (c) DEV3 and BTAP swaps their role by master-slave switch command. (d) DEV3 exchanges data with BTSA. (e) DEV3 closes communication channel.

Results and Discussion

The authors developed and commercialized a Bluetooth-based sensor network system. The system provides a virtual serial port for a server to connect with Bluetooth devices over WiFi and Ethernet as shown in Figure 3. Figure 4(a) shows the developed BTAP (Takebishi Corp.).

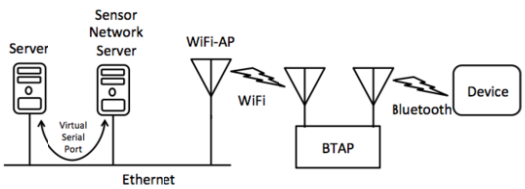


Figure 3 – Diagram of the developed Bluetooth-based sensor network system

(a) (b)

Figure 4 – The devices introduced to the hospital. (a) BTAP antenna (Takebishi Co.). (b) Barcode Reader (KoamTac Inc.).

Kyoto University Hospital [4] introduced the developed Bluetooth-based sensor network system with approximately 400 BTAP in 2011. The main target device was a Bluetooth-based barcode reader (KoamTac Inc. KDC 300, Figure 4(b)) for auto-ID / barcode enabled medication administration (ABMA) system [5]. Kyoto University Hospital introduced 1500 barcode readers. Each nurse and resident has a barcode reader for performing approximately 2700 safety checkups for transfusions and injections per day. Since January 2011, when the system was put into operation, the hospital has not experienced any major issues caused by lost communication, including roaming trouble, to the moment that this paper was written (April 2015).

The authors investigated all the comunication errors of the ABMA system on a standard business day, Thursday, October 9th, 2014. 2260 checkups were performed on the ABMA, and 32 communication errors occurred.

Figure 5 shows the basic communication protocol of the ABMA system. As the connection should be always initiated by the barcode device, the atomic communication of the system always consists of four messages, that is, barcode data, ack from server, response from server, and ack from the reader. During this atomic communication, no barcode device can change communicating BTAP.

As the ABMA checkup process consists of two atomic communications as shown in Figure 5, eight packets are needed to be sent to fulfill safety checkup. If any single packet within an atomic communication is lost, the ABMA system asks the clinical staff to re-read the barcode again. Therefore, the error rate of the observed day was $32 \div ((2260 + 32) \times 8) \times 100 = 0.17\%$.

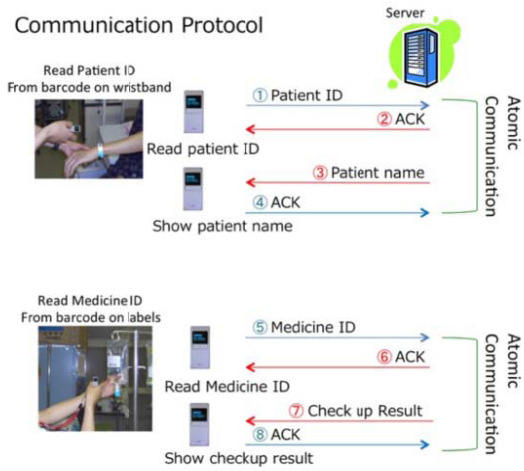


Figure 5 – Communication Protocol of the ABMA system

Additionally, packet can be lost during any connection between the device and the sensor network server. As shown in Figure 3, the communication is mediated by Ethernet and WiFi as well as Bluetooth. Thus, it is not the case that all observed error was caused by the roaming error of the developed system.

Thus, the developed roaming system worked quite well in the real clinical setup of Kyoto University Hospital, and provides clear efficiency for nurses to enable with their daily safety checkups with single tiny device. As a matter of fact, the average number of performed checkups per day increased from 1825 (January 2011) to 2260 (December 2014) after the introduction of the Bluetooth-based ABMA system. The number increased eight percent (from 1825 to 1973) in the first five months. This quite steady increase tells that the reliability of the ABMA system, supported by the realiability of the proposed Bluetooth roaming, earning the clinical staff trust. This fact clearly confirms the effectiveness of the proposed roaming mechanism.

As the proposed method realized by slight modification of the Bluetooth protocol, slight customization of firmware enables any Bluetooth-equipped devices, including the barcode reader introduced for Kyoto Univrsity Hospital, to connect to the developed sensor network.

The proposed protocol assumes that the connecting sensors send data intermittently. As the most of the vital sensing devices used in a standard hospital ward, such as blood pressure meter, body temperature meter, and weight meter, produce small size packets intermittently, just as the ABMA system implemented on the proposed sensor network that exchanges four short messages and four ACK packets. Therefore, the Bluetooth-based sensor network based on the proposed roaming mechanism may be able to accept most of such sensor devices.

On the other hand, in order to make the sensor network be enabled for continuous sensing devices, such as electrocardiogram (EKG) or oxygen saturation monitor (SpO2), certain compensation mechanism, such as buffering at the device side, should be considered.

Conclusion and Outlook

This paper proposed and implemented a method to provide roaming functionality for Bluetooth to realize Bluetooth-based sensor network for healthcare industries. The introduction of the developed Bluetooth-based sensor network system for daily activities in a university hospital with approximately 1000 admitted patients per day for five years clearly confirms the effectiveness and the availability of the proposed approach.

As the Bluetooth becomes a quite common wireless interface for medical devices, the Bluetooth-based sensor network may enable HISs to equip various medical and non-medical devices. Consequently, the sensor network expands capability of information communication technology to advance clinical services.

Acknowledgments

The research is partly funded by the Ministry of Internal Affairs and Communication of Japan, and JSPS Grant-in-Aid for Scientific Research (B) number 2580106. The authors would like to thank Takebishi Corp., Shimadzu System Development Corp. and IBM Japan for their precious support in developing, commercializing and implementing the developed sensor network and ABMA system.

References

[1] Kuroda, T.; Sasaki, H.; Suenaga, T.; Masuda, T.; Yasumuro, Y.; Hori, K.; Ohboshi, N.; Takemura, T.; Chihara, K.; Yoshihara, H. Embedded Ubiquitous Services on Hospital Information Systems. IEEE Transactions on Information Technology in Biomedicine 2012, 16(6), 1216-1223.

[2] Kuroda, T.; Nakashima, N.; Kaneda, M.; Sato, K.; Ohboshi, N.; Okada, A.; Okamoto, K.; Yoshihara, H.; Akiyama, T. Evaluating the Workload Reduction of Automatic Vital Data Transmission, Advanced Biomedical Engineering 2013, 2, 124-129.

[3] Caroll, R.; Cnossen, R.; Schnell, M.; Simons, D. Continua: An interoperable personal healthcare eco- system. IEEE Pervasive Computing 2007, 6 (4), 90–94.

[4] Kyoto University Hospital Guidance. Available online: http://www.kuhp.kyoto-u.ac.jp/english/outline/index.html (Accessed Oct. 13, 2014).

[5] Wright, A.A. Bar coding for patient safety. New England Journal of Medicine 2005, 353, 329-331.

Address for correspondence

Tomohiro Kuroda, Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Shogo-in Kawahara-cho 54, Sakyo-ku, Kyoto, Japan.
Email: tomo@kuhp.kyoto-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-202

On Building an Ontological Knowledge Base for Managing Patient Safety Events

Chen Liang, Yang Gong

School of Biomedical Informatics, University of Texas Health Science Center at Houston, TX, USA

Abstract

Over the past decade, improving healthcare quality and safety through patient safety event reporting systems has drawn much attention. Unfortunately, such systems are suffering from low data quality, inefficient data entry and ineffective information retrieval. For improving the systems, we develop a semantic web ontology based on the WHO International Classification for Patient Safety (ICPS) and AHRQ Common Formats for patient safety event reporting. The ontology holds potential in enhancing knowledge management and information retrieval, as well as providing flexible data entry and case analysis for both reporters and reviewers of patient safety events. In this paper, we detailed our efforts in data acquisition, transformation, implementation and initial evaluation of the ontology.

Keywords:

Medical error; Patient safety; Ontology; Knowledge management

Introduction

The increasing high rate of medical errors indicates that patient safety is a prominent issue [1, 2]. A recent study reported that preventable medical errors cause annual deaths of 210,000 to 440,000 in the United States [3]. The magnitude of these medical errors, as well as near misses and unsafe conditions, has raised public awareness on patient safety and interest in research. For the purpose of preventing error occurrence, much attention has been drawn into reasoning about systemic factors that contribute to the errors, while the road block appears to be the disclosure of patient safety events [1]. It is documented that the major obstacle to disclosing patient safety events and proposing systematic solutions is due to limited functionalities of a reporting system [4]. The core functionality of a reporting system is thought to be collecting, analyzing and learning from the existing mistakes [5-7]. Therefore, the systems should include functional modules such as data acquisition, knowledge management, information retrieval, and beyond.

Reporting systems are pervasively used in the United States [4, 8], Australia [9], United Kingdom [10] and other countries. Nevertheless, debates on the effectiveness of such systems still remain. So far , there has hardly been any study that reports a decrease of medical errors, mortality or morbidity directly due to the intervention of a reporting system [4]. Obviously, more efforts are needed to prove the value of the reporting systems. A first-line question is why the reporting systems are unsuccessful in demonstrating the value. Among various reasons, data quality is thought to be a major concern. Patient safety reports containing detailed narratives are helpful in replicating the incidents and further translating into improved patient safety. The narratives are different from other data types stored in generic health information systems, such as electronic medical record (EMR) data and clinical notes. Reporting systems collect a variety of data ranging from structured to unstructured formats and therefore may cause a data inconsistency issue, which leads to incompleteness and inaccuracy of the data entry [11]. For example, a reporting system with a structured data entry can oftentimes force the reporters to choose “other or miscellaneous” when they are asked to categorize a patient safety event [12]. In contrast, unstructured data (free text) are unconstrained in offering detailed information but they are not immune from issues. An obvious pitfall for the use of unstructured data in the reporting system is time efficiency. Many reporters are working under time pressure or in a multi-task mode and thus they may not have sufficient time to provide a complete and detailed report to the systems [11]. A recent study using a text prediction method intended to mitigate the abovementioned problems [13]. With this method, the system provides prompting information (suggested words/phrase to use) to the reporters at the time when they are typing free text into the data entry portal. The text prediction technique provides insights into extracting and organizing the semantic information based on the free text with the only limitation being that the prediction list was manually prepared by domain experts. Another pitfall for the use of unstructured data appears in data processing. Similar to clinical notes, typos, abbreviations, and nonstandard acronyms are typically intertwined with free text data. These have become barriers for data pre-processing, such as de-identification, and classification by natural language processing (NLP), and thus cost extra time for reviewers in understanding the data.

Taxonomies have been used to address these problems as they intend to manage patient safety events as a knowledge base. The taxonomies used to document and classify patient safety reports can be traced back to 1987 when the Australian Patient Safety Foundation (APSF) originally reported the Australian Incident Monitoring System [14]. Later on, a series of well known taxonomies were put into use, which include the JCAHO patient safety event taxonomy [15], the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)’s taxonomy of medication errors [16], the Neonatal Intensive Care system (NIC) [17], the Pediatric Patient Safety taxonomy (PED) [18], the Preliminary Taxonomy of medical errors in Family Practice (PTFP) [19], the Taxonomy of Nursing Errors (TNE) [20], and the Adverse Event Reporting Ontology (AERO) [21]. While these taxonomies served primarily as domain specific knowledge bases, the rapid increase of patient safety data calls for a sharable knowledge base organized by a unified language system.

In sum, more efforts are needed for improving data quality, sharing and learning from patient safety events across the individual systems. A significant challenge to increase data

quality remains in the development of a unified domain knowledge base and the effective use of the data. To explore solutions to the problems, we aim to build a semantic web ontology (Medeon) using W3C open standard Web Ontology Language (OWL). The Medeon serves as a unified knowledge base for organizing patient safety events, and as a supporting component that facilitates the user end applications towards enhancing data entry and data quality.

Materials and Methods

Semantic Web Ontology

A taxonomy as a controlled vocabulary in hierarchy has been used in patient safety reporting for years. Ontologies are explicit specifications of conceptualized definitions and relationships where these specifications define a taxonomy of the knowledge [22]. Specifically, an ontology models the real-world knowledge by encoding the entities and the relationships among the entities. We aimed to develop an ontology to replace the role of a taxonomy because the ontology can provide a broader application over taxonomies. We chose OWL and semantic web technologies because they jointly provide a unique advantage for machine understandable semantics and descriptive logic reasoning. OWL has an advantage that allows us to identify unique patient safety terminologies or concepts that may appear under different names or originate from different sources. This advantage largely reduces the ambiguity in medical terminologies, and may advance the knowledge management of patient safety events.

Previous work on building ontologies for patient safety events employed various methodologies (i.e., techniques, tools, procedures and guidelines) [23-25], yet these approaches are lacking computer understandable representations. In the present work, we refer to the *Semantic Web for the Working Ontologist* for theories and general guidelines commonly employed in developing OWL ontologies [26]. Protégé (V4.3.0) was employed to implement the ontology.

learning across the reporting systems through a unified language serving as a common denominator. With this view, the Common Formats (v1.2) developed by the Agency for Healthcare and Research Quality (AHRQ) were employed as the taxonomy where we extracted and encoded semantic knowledge into Medeon. Recognized as a unified standard of reporting patient safety events, the Common Formats are designed to specify and collect event information, which range from general concerns to frequent occurrences and serious types of events. We borrowed the hierarchical structure in the Common Formats to build the OWL classes and rephrased the narrative data in the Common Formats to construct OWL incidents and objective properties. At the top level, four OWL classes consist of *Circumstances_of_Event, Patient_Information,Reporting_Reporter_and_Report_Information,* and *Type_of_Event*, with the maximum depth of four. Figure 1 shows the visualization of expanded classes. The OWL Object Property was defined as IsA property, for example, *Inattention IsA Human_factors*.

Data Acquisition

At an initial stage, all the entities and relationships implemented in Medeon were extracted from the Common Formats. Healthcare event reporting form (HERF), patient information form (PIF), and summary of initial report (SIR) in the Common Formats are regarded as a comprehensive and relatively complete collections of entities that can represent patient safety events. Therefore, a direct translation was performed in order to encode those entities from the Common Formats to Medeon. To obtain high quality data in the Common Formats, we followed a set of principles as guidelines [27]. Table 1 provides a brief description on the principles. We borrowed eight principles that were separated into three dimensions to guide the rephrasing of the language used in the ontology. Note that we did not include 'social quality' from the Dimensions in the original literature since this dimension measures the ontology comparing it to the existing ontologies and emphasizes the utility of the ontologies which are not applicable to the project. When the

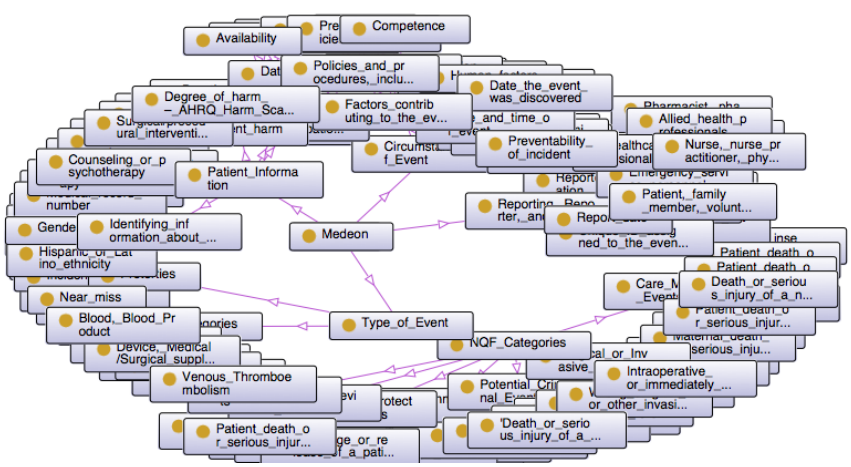


Figure 1 - OWL Class Visualization by OntoGraf (V1.0.1) in Protégé (V4.3.0).

The Meta Ontology

Patient safety taxonomies have been used as a reference where hierarchies, entities, and relationships can be used as candidate materials [23]. While these taxonomies served primarily as standards of domain specific taxonomies, the rapid growth in medical information needs a knowledge base for sharing and

translation of entities was completed, we imported those entities into Medeon via Protégé. Data consistency was checked through Protégé build-in modules to ensure that no logical conflict existed in the ontology.

Evaluation

The evaluation is intended to provide a comprehensive report pertaining to the effectiveness and validity of ontology in multi-dimensions. Our evaluation design included the assessment of the ontology itself and the user experience. We designed a questionnaire with a 5-point Likert scale to collect the measurable data from domain experts interested in using the ontology in their daily work.

Below we enumerate a set of questions in the questionnaire. Each question is equipped with answers to a 5-point Likert scale (i.e., 1=very disappointed; 2=disappointed; 3=neutral; 4=good; 5=very good).

1. The phrases used in the vocabulary are well formed and the words are well arranged.
2. The terms used in the vocabulary can explain the meanings of real-world concepts.
3. The terms that appear in the vocabulary are clear.
4. The vocabulary represents the designated domain and provides sufficient knowledge to the user.
5. The claims the vocabulary makes are reasonable.
6. The vocabulary can satisfy your requirements when you use it to categorize the case you are reviewing.
7. Please rate the overall satisfaction based on your experience using the vocabulary.

To make sure that the questionnaire reached the confidence level on effectiveness and validity, we employed a pre-measurement to assess the content-validity and inter-rater reliability to guide the final revision of the questionnaire. The content-validity measures to what extent the designed questions subjectively reflect the tasks they purport to measure. The inter-rater reliability measures the degree of agreement among raters. Three domain experts used the pre-measurement to validate the questionnaire where randomly selected patient safety reports were provided in the task.

The questions listed below were used for measuring content-validity. Each question was instructed to be answered on a 4-point scale (i.e., Not relevant; Somewhat relevant; Quite relevant; Highly relevant)

1. "The phrases used in the vocabulary are well-formed and the words are well-arranged."
   Does the scale purport to measure "The correctness of syntax."?
2. "The terms used in the vocabulary can explain the meanings of real-world concepts."
   Does the scale purport to measure "The meaningfulness of terms."?
3. "The terms that appear in the vocabulary are clear."
   Does the scale purport to measure "The clarity of terms."?
4. "The vocabulary represents the designated domain and provides sufficient knowledge to the user."
   Does the scale purport to measure "The comprehensiveness of the vocabulary in a certain domain."?
5. "The claims the vocabulary makes are reasonable."
   Does the scale purport to measure "The accuracy of information."?
6. "The vocabulary can satisfy your requirements when you use it to categorize the case you are reviewing."
   Does the scale purport to measure "Whether the vocabulary specifies agent's specific requirements."?
7. "Please rate the overall satisfaction based on your experience using the vocabulary."
8. Does the scale purport to measure "The overall satisfaction to the vocabulary."?

Results

Upon the completion of Medeon, we obtained a semantic web ontology in OWL format. The ontology represented, with necessary conceptualization and translation, the entities and relationships of patient safety knowledge that were extracted from the Common Formats. The ontology was constructed in a hierarchy with four top-level classes where each contained sub-classes with a maximum depth of four levels. An example of OWL individuals is shown in Figure 2. As in the preliminary stage, these individuals may be incomplete yet they represent the most frequently used concepts and terminologies appearing in the Common Formats. Note that OWL classes and individuals, as well as the OWL properties, are open to expand. That being said, the knowledge we borrowed from the Common Formats serve as building blocks for further development without limiting patient safety ontology.

Two domain experts participated in the pre-measurement. The results showed a 100% agreement for the inter-rater reliability and 100% for content validity.

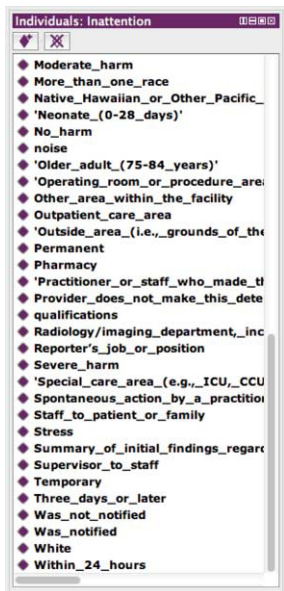


Figure 2 - OWL Individuals in Protégé (V 4.3.0).

Discussion and Future Work

The ontology primarily serves as a knowledge base to model the taxonomies broadly used for patient safety reports. With this role, the proposed ontological approach aims to meet the challenges in the development of reporting systems. Among the many factors fundamental for a successful reporting system, data quality has been a major concern. An outstanding reporting system should be able to collect quality data that link to the procedures and factors threatening patient safety in a timely manner. Nevertheless, a great number of reporting systems are suffering from low quality data due to inefficiency and ineffectiveness of data entry [11, 28, 29]. To improve the data quality, much effort has been made in increasing the number of reports, but the increase in quantity does not improve the system performance since the very crux of the

problem remains in knowledge management. In fact, the reporting systems in use generate a great volume of patient safety reports, which on the contrary are becoming a burden for data processing.

The ontology is what we believe is a suitable approach in patient safety reporting. Given a patient safety report, oftentimes it needs to be labeled with multiple categories in a hierarchical knowledge base. Neither plain reporting forms nor patient safety taxonomies can easily solve this problem during the submission of a report to the system or retrieving a report from the system. For example, a case being labeled under 'lighting' may be also labeled under 'contributing factor', 'environment', and 'patient fall', assuming that 'contributing factor' and 'environment' are the super classes of 'lighting', while 'patient fall' is under the other super class yet has certain associations to 'lighting'. It could be time consuming when reporting such a 'lighting' case in plain reporting forms and it could be difficult to extract the association between 'lighting' and 'patient fall', since they belong to different classes or different hierarchical levels. Fortunately, an ontology can contribute to these aspects since it encodes all the relationships among entities and is able to query via path expression.

Ontology provides insights into an efficient and user-friendly data entry framework. Researchers have found close associations between data entry and the system performance in terms of the completeness and accuracy of patient safety reports [30, 31]. The majority of patient safety reports are recorded in free text. Although the free text might be an efficient and natural means for users to deliver an informative case, it could be costly to turn the raw information into a cognitively organized and manageable format for professionals to utilize . Applying semantic web ontology to organize the text data can support effective data entry. Such a use case includes web portals where ontology is used for defining terminologies and concepts (meanings) for an area of knowledge. For example, 'patient fall', 'female', 'slip', and 'emergency room' are terminologies that can be found in the patient safety ontology. The ontology can also define a concept like 'All female slips in emergency room are considered as patient falls.' This provides an interesting angle to look into; the latent yet important information in the patient safety reports since a few distributed semantics (terminologies and relationships) appear to be sufficient for representing the most significant meanings in a report.

To our best knowledge, the use of AHRQ Common Formats as the source taxonomy to build semantic web ontology is an initiative. Patient safety reporting systems never lack taxonomies, but a universal and computer understandable knowledge base. The Common Formats provide a compatible format that includes patient safety events ranging from a series of general concerns, frequent occurrences and serious types of events andmore importantly, the implementation of the Common Formats in an ontological framework offers an open environment to aggregate and share the patient safety knowledge base by cooperating with other data sources and ontologies.

Our design and implementation have challenges and limitations. When building the ontology, it is most challenging in mapping between discrepant data sources due to the distinction among existing taxonomies in terms of the hierarchical structures and synonymous terminologies. We envision the use of NLP techniques and automatic classifier (i.e., k nearest neighbor) could facilitate the process, as we will expand Medeon in the next step. Also, a view on a unified coding system, such as Unified Medical Language System (UMLS), is definitely helpful. On the other hand, debates regarding validity and effectiveness are always in ontological studies Therefore, we must continue the evaluation study of the present ontology.

Table 1 - Design Principles for High Quality Ontologies.

Dimensions	Attributes	Description
Syntactic quality	Lawfulness	Correctness of syntax
	Richness	Breadth of syntax used
	Interpretability	Meaningfulness of terms
Semantic quality	Consistency	Consistency of meaning of terms
	Clarity	Average number of word senses
	Comprehensiveness	Number of classes and properties
Pragmatic quality	Accuracy	Accuracy of information
	Relevance	Relevance of information for a task

Conclusion

The development of a knowledge base for patient safety reporting systems is imperative for both practice and research. With the aim of establishing a comprehensive knowledge base, we employed a semantic web ontology that plays a key role underlying the reporting systems. The present ontology built on the Common Formats serves as the building blocks towards a unified knowledge base, with which the reporting systems are expected to support comprehensive data entry and increase the data quality. We envision that utilizing a semantic web ontology would facilitate information retrieval and reuse of the narrative data for expert review, clinical decision-making and education. Moving forward, the swift growth in aggregate data requires a sustainable knowledge base to keep abreast with the latest reporting events. Our design must be open-minded to glean knowledge from the most recently reported data and tremendous amount of historical data.

Acknowledgments

This project is supported by a grant from AHRQ, 1R01HS022895 and a patient safety grant from the University of Texas system. We thank Dr. Jing Wang for her invaluable comments in designing the evaluation questionnaires for the ontology.

References

[1] Kalra J, Kalra N, Baniak N. Medical error, disclosure and patient safety: A global view of quality care. Clinical biochemistry. 2013;46(13):1161-9.

[2] Kalra J. Medical errors: an introduction to concepts. Clinical biochemistry. 2004;37(12):1043-51.

[3] Joshi MS, Anderson JF, Marwaha S. A systems approach to improving error reporting. Journal of healthcare information management: JHIM. 2001;16(1):40-5.

[4] Woodward HI, Mytton OT, Lemer C, Yardley IE, Ellis BM, Rutter PD, et al. What have we learned about interventions to reduce medical errors? Annual review of public health. 2010;31:479-97.

[5] Leape LL, Abookire S, World Health O. WHO draft guidelines for adverse event reporting and learning systems: from information to action: World Health Organization; 2005.

[6] Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman EJ, et al. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. Quality and Safety in Health Care. 2007;16(1):40-4.

[7] Weissman JS, Schneider EC, Weingart SN, Epstein AM, David-Kasdan J, Feibelmann S, et al. Comparing patient-reported hospital adverse events with medical record review: do patients know something that hospitals do not? Annals of Internal Medicine. 2008;149(2):100-8.

[8] Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System: National Academies Press; 2000.

[9] Runciman WB. Lessons from the Australian Patient Safety Foundation: setting up a national patient safety surveillance system—is this the right model? Quality and Safety in Health Care. 2002;11(3):246-51.

[10] Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Quarterly. 2001;79(2):281-315.

[11] Gong Y. Data consistency in a voluntary medical incident reporting system. Journal of medical systems. 2011;35(4):609-15.

[12] Gong Y, Richardson J, Zhijian L, Alafaireet P, Yoo I, editors. Analyzing voluntary medical incident reports. AMIA Annual Symposium proceedings/AMIA Symposium AMIA Symposium; 2007.

[13] Hua L, Wang S, Gong Y. Text Prediction on Structured Data Entry in Healthcare. Methods Inf Med. 2008;47(1):8-13.

[14] Spigelman AD, Swan J. Review of the Australian incident monitoring system. ANZ journal of surgery. 2005;75(8):657-61.

[15] Chang A, Schyve PM, Croteau RJ, O'Leary DS, Loeb JM. The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. International Journal for Quality in Health Care. 2005;17(2):95-105.

[16] Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. Journal of biomedical informatics. 2004;37(3):193-204.

[17] Suresh G, Horbar JD, Plsek P, Gray J, Edwards WH, Shiono PH, et al. Voluntary anonymous reporting of medical errors for neonatal intensive care. Pediatrics. 2004;113(6):1609-18.

[18] Woods DM, Johnson J, Holl JL, Mehra M, Thomas EJ, Ogata ES, et al. Anatomy of a patient safety event: a pediatric patient safety taxonomy. Quality and Safety in Health Care. 2005;14(6):422-7.

[19] Dovey SM, Meyers DS, Phillips RL, Green LA, Fryer GE, Galliher JM, et al. A preliminary taxonomy of medical errors in family practice. Quality and Safety in Health Care. 2002;11(3):233-8.

[20] Woods A, Doan-Johnson S. Toward a Taxonomy of Nursing Practice Errors. Nursing Management. 2002;33(10):45-8.

[21] Courtot M, Brinkman RR, Ruttenberg A. The Logic of Surveillance Guidelines: An Analysis of Vaccine Adverse Event Reports from an Ontological Perspective. PloS one. 2014;9(3):e92632.

[22] O'Leary DE. Using AI in knowledge management: Knowledge bases and ontologies. IEEE Intelligent systems. 1998;13(3):34-9.

[23] Mokkarala P, Brixey J, Johnson TR, Patel VL, Zhang J, Turley JP. Development of comprehensive medical error ontology. AHRQ: Advances in Patient Safety: New Directions and Alternative Approaches invited paper, under review. 2008.

[24] Gong Y, Zhu M, Li J, Turley J, Zhang J, editors. Clinical communication ontology for medical errors. AMIA Annual Symposium Proceedings; 2006: American Medical Informatics Association.

[25] Stetson PD, McKnight LK, Bakken S, Curran C, Kubose TT, Cimino JJ, editors. Development of an ontology to model medical errors, information needs, and the clinical communication space. Proceedings of the AMIA Symposium; 2001: American Medical Informatics Association.

[26] Allemang D, Hendler J. Semantic web for the working ontologist: effective modeling in RDFS and OWL: Elsevier; 2011.

[27] Burton-Jones A, Storey VC, Sugumaran V, Ahluwalia P. A semiotic metrics suite for assessing the quality of ontologies. Data & Knowledge Engineering. 2005;55(1):84-102.

[28] Pronovost PJ, Morlock LL, Sexton JB, Miller MR, Holzmueller CG, Thompson DA, et al. Improving the value of patient safety reporting systems. Advances in patient safety: new directions and alternative approaches. 2008;1.

[29] Gong Y, Terminology in a voluntary medical incident reporting system: a human-centered perspective. Proceedings of the 1st ACM International Health Informatics Symposium; 2010: ACM.

[30] Johnson CW. Failure in safety-critical systems: a handbook of accident and incident reporting. University of Glasgow Press: Glasgow, Scotland; 2003.

[31] Barach P, Small SD. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. Bmj. 2000;320(7237):759-63.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St, Suite 600

Houston, TX, 77030, USA

Email: Yang.Gong@uth.tmc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-207

Case Study: Applying OpenEHR Archetypes to a Clinical Data Repository in a Chinese Hospital

Lingtong Min[a], Li Wang[a], Xudong Lu[a*] and Huilong Duan[a]

[a] *College of Biomedical Engineering & Instrument Science, Zhejiang University, China*

Abstract

openEHR is a flexible and scalable modeling methodology for clinical information and has been widely adopted in Europe and Australia. Due to the reasons of differences in clinical process and management, there are few research projects involving openEHR in China. To investigate the feasibility of openEHR methodology for clinical information modelling in China, this paper carries out a case study to apply openEHR archetypes to Clinical Data Repository (CDR) in a Chinese hospital. The results show that a set of 26 archetypes are found to cover all the concepts used in the CDR. Of all these, 9 (34.6%) are reused without change, 10 are modified and/or extended, and 7 are newly defined. The reasons for modification, extension and newly definition have been discussed, including granularity of archetype, metadata-level versus data-level modelling, and the representation of relationships between archetypes.

Keywords:

Archetype, openEHR, Clinical Data Repository, Chinese hospital.

Introduction

A Clinical data repository (CDR) is a repository that stores the clinical data integrated from various kinds of clinical information systems for analysis and research. An increasingly larger demand for clinical research and application has drawn international initiatives' attention and propels institutions such as the Mayo Clinic [1], Intermountain Healthcare [2], Stanford Medical Center [3], Massachusetts General Hospital [4] and Columbia University Medical Center [5], to build CDRs.

As healthcare is a highly complicated and rapidly developing domain, the flexibility of the data model in CDR is rather indispensable. Being an open architecture for Electronic Health Record, openEHR advocates a dual-level methodology to conduct data modelling for clinical information [6]. It also provides a flexible modeling methodology to adapt to the evolution of clinical concept and knowledge. The architecture of openEHR includes reference model and archetype to separate clinical knowledge from clinical information. The reference model defines the data type, data structure, basic framework of EHR and represents the global characteristics of health record entries [7]. The archetype is a conceptual model, which is built on the base of clinical knowledge by clinical experts and specifies constraints on the reference model.

Although openEHR methodology has been widely carried out and implemented in Europe and Australia [8-10], only a few researches studies utilizing openEHR have been conducted in China. Most of the existing publications in China were limited to the introduction or translation of the basic principles and concepts of openEHR [11-13]. Only few were related to openEHR implementation in the area of research [14], while none of them focuses on data modeling in clinical settings. The possible obstacles are the differences in clinical process of hospital and the healthcare management mechanisms between countries. To implement openEHR in China, an investigation of the feasibility of using openEHR methodology to model clinical data in China is a necessity. This paper introduces a case study of applying the openEHR archetypes to CDR in a Chinese hospital.

Methods

The selected tertiary class A hospital in the case study has deployed several information systems including HMIS (hospital management information system), CIS (clinical information system), PACS (picture archiving and communication system), RIS (radiology information system), and LIS (laboratory information system). The clinical data has been scattered in these silo systems, and the physicians have to access different patient data in corresponding system respectively. It is crucial to build a CDR to integrate all the clinical data from these heterogeneous information systems and provide real-time data access services for applications to browse the complete set of patient data in data viewing applications. To achieve this target, the CDR needs to contain all essential clinical data including domains of patient demographics, encounters, medication, imaging examination, and laboratory test. The authors have designed a 6-step method to model the data in CDR via openEHR.

1. Analyze MOH standards

The data requirements of the CDR are first analyzed with regard to several standards. Chinese MOH (Ministry of Health) has published a set of standards related to healthcare data sets to facilitate information collection, storage, and exchange, such as "WS 363-2011 Health data element dictionary" (WS 363-2011) and "WS 445-2014 Basic dataset of electronic medical record" (WS 445-2014). "WS 363-2011" defines the identification, naming, meaning, data type, representation, and value set for all the data elements in healthcare domain. The purpose of this standard is to provide the standardized definition of data elements for all applications in healthcare domain. "WS 445-2014" specifies the typical business model and clinical documentation of the electronic medical record (EMR) together with the data sets under this architecture. The data elements used in this

standard conform to "WS 363-2011", except for some that are further constrained to adapt to the EMR context.

Firstly, the standards of "WS 445-2014" and "WS 363-2011" were analyzed according to the requirements of the CDR scope, content, and data elements. The coverage of standards for the requirements is listed in Table 1 and Table 2.

To be more specific, for each CDR data requirement, we need to find the corresponding data sets of "WS 445-2014" and refer to "WS 363-2011" for collecting the necessary data items in CDR, "none" indicates that the CDR requirements are not included in standards. For instance, the imaging examination requirements in the CDR refers to three data sets of "outpatient and emergency medical record", "examination and laboratory test record" and "inpatient progress note" in "WS 445-2014". After that, the relevant imaging examination data items can be found within "identification", "assistant examination" and "healthcare organization" in "WS 363-2011".

Table 1 – CDR requirements and WS 445-2014

WS 445-2014	CDR requirements
1) medical record summary	patient demographics, encounters
2) outpatient and emergency medical record	imaging examination
3) outpatient and emergency prescription	medication
4) examination and laboratory test record	imaging examination, laboratory test
5) general therapy and treatment record	medication
6) delivery record of therapy and treatment	none
7) nursing operation records	none
8) nursing valuation and plan	none
9) informing information	none
10) home page of inpatient medical record	none
11) home page of inpatient medical record summary of TCM	none
12) admission record	encounters
13) inpatient progress note	imaging examination
14) inpatient order	medication
15) discharged brief	encounters
16) transfer record	encounters
17) medical institution information	none

Table 2 – CDR requirements and WS 363-2011

WS 363-2011	CDR requirements
1) identification	patient demographics, encounters, medication, imaging examination, laboratory test
2) demographics and social economics characteristics	patient demographics
3) health history	none
4) health risk factor	none
5) chief complaint and symptom	none
6) physical examination	none
7) assistant examination	imaging examination
8) laboratory examination	patient demographics, laboratory test
9) diagnosis	encounters
10) medical assessment	encounters
11) medical plan and intervention	encounters, medication
12) health expenditure	none
13) healthcare organization	patient demographics, encounters, medication, imaging examination, laboratory test
14) health personnel	none
15) drug and material	medication
16) health management	none

2. Analyze existing database schemas

The data items defined in the standards can only cover part of the CDR data requirements, and the schemas of the existing heterogeneous information systems should also be taken as a source for collecting the necessary data elements in CDR as shown in Table 3. For example, the data items of imaging examination can be found from Exam request, Exam item, Exam procedure, Image, and Report four tables in the existing database schemas.

Table 3 – CDR requirements and existing database schemas

CDR requirements	Information systems	Table schemas
patient demographics	HMIS	Patient
encounters	CIS	Visit, Inpatient admission
imaging examination	RIS	Exam request, Exam item, Exam procedure, Image, Report
laboratory test	LIS	Test request, Test item, Report, Sample
medication	CIS	Medication order

3. Merge data items

Data items collected from the above two steps are merged into the final data set for CDR. If the data type, value set, or coding for same data element is not compatible in the standards and the database of existing systems, the one from the standards was used. After merging, 217 data elements were finally defined in a structured format and ready to be used as shown in Table 4. The data items of imaging examination is composed of 38 items among which 13 items came from standards and 25 items came from information systems.

Table 4 – Number of data items collected from standards and information systems

CDR requirements	Standards	Information systems
patient demographics	18	21
encounters	16	45
imaging examination	13	25
laboratory test	10	17
medication	16	36

4. Organize data items into concepts

Since each archetype models only one distinct concept, the data items should be organized into concepts before modeling. Although the data items are classified into basic data sets in the standards, the classification is mainly clinical documentation oriented and the data sets often contain a number of distinct concepts. For example, the data sets in "part 4) examination and laboratory test record" of "WS 445-

2014" contain two sub-domains of laboratory test and examination, and there are concepts about request, result, report, and specimen in each sub-domain. The data elements in "WS 363-2011" are just listed in a row one by one. On the other hand, the data schemas from different information systems are heterogeneous and overlapped. To organize the data items from both standards and data schemas of different systems, an entity-relationship concept modeling process is carried out and results in a group of 26 concepts. Semantic overlapping between concepts is avoided to our best to comply with the single archetype for single concept principle. The derived clinical concepts are shown in Figure 1.

5. Map concepts to archetypes

To achieve maximum reusability, the public archetype repository Clinical Knowledge Manager (CKM) is searched for matching archetypes with each derived concept with the name or other key words. The candidate archetypes filtered by searching are analyzed in details. If all the data items of the clinical concept are covered by the existing archetype, this archetype can be directly adopted without any modification. If only part of the data items of clinical concept is covered by the existing archetype, this archetype needs to be modified and extended. If more than one matching archetypes exist, the most semantically suited one was selected and then checked for whether it could completely cover or partially cover the concept. If there is no matching archetype for the clinical concept, a new one is defined. For archetype modification and extension, several operations are illustrated in the openEHR specification [15] such as revision, specialization, and new version.

6. Model relationships between concepts

After mapping concepts to archetypes for the CDR, the relationships between the concepts also need to be modeled. There are two general methods to express relationships between archetypes, archetype slot and link. Archetypes can be composed to express valid possibilities for larger structures of data from different levels of ontological hierarchy of the reference model. Such compositional connections are termed as 'slots'. For example, "Section" and "Entry" archetypes can be composed through slots to represent the structure similar to clinical document with head, body, and content. Currently, almost all the archetypes in CKM use slots to express relationships between archetypes. "Link" is an attribute root in the deep architecture of the openEHR reference model and can refer to many other archetype structures.

Results

26 archetypes are identified to cover all the data requirements of the CDR, shown in Figure 2. 9 archetypes, listed in Table 5, are from CKM and can be directly reused without any modification, which account for 34.6% of all 26 archetypes. 10 archetypes, listed in Table 6, are modified and extended, among which 2 are revisions by replacing compatible data types with some data items, 8 are modified by adding data items, no specialization and no new version archetypes. 7 archetypes, listed in Table 7, are newly defined since not all the clinical concepts are covered by the existing archetypes, which include 2 for laboratory test, 2 for examination, 2 for medication orders, and 1 for transfer management.

So far, the 26 archetypes have been used to implement the CDR based on the relational database that provides data access service in the hospital. Clinical data has been integrated from the heterogeneous systems into the CDR and a clinical data viewer for clinicians has been developed and used in the clinical practice.

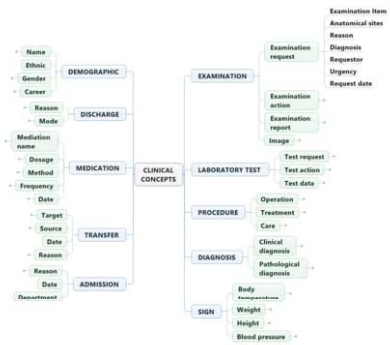


Figure 1 – The clinical concepts of the CDR

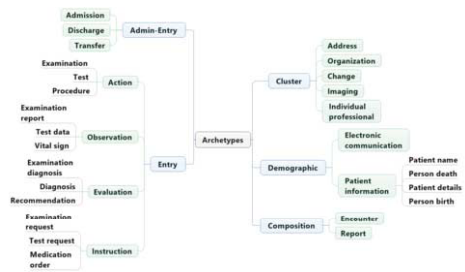


Figure 2 – The archetype structure of the CDR

Table 5 – Not changed archetypes

Archetype
ADMIN_ENTRY.discharge_admin_info.v3
CLUSTER.address.v1
CLUSTER.lab_result_annotation.v1
CLUSTER.medication_amount.v1
CLUSTER.organisation.v1
CLUSTER.specimen.v1
DEMOGRAPHIC-ADDRESS.electronic_communication.v1
DEMOGRAPHIC-PARTY_IDENTITY.person_name.v1
DEMOGRAPHIC-PERSON.person-patient.v1

Table 6 – Modified and extended archetypes

Archetype
DEMOGRAPHIC-CLUSTER.person_identifier.v1.1
DEMOGRAPHIC-ITEM_TREE.person_details.v1
ACTION.imaging_exam.v1.1
ACTION.medication.v1.1
CLUSTER.medication_admin.v1
ADMIN_ENTRY.admission.v1.1
COMPOSITION.encounter.v1.1
INSTRUCTION.medication.v1.1
INSTRUCTION.request-imaging_exam.v1.1
INSTRUCTION.request-lab_test.v1.1

Table 7 – Newly developed archetypes

Archetype
ACTION.lab_test.v1
ACTION.medication_order_schedule.v1
ADMIN_ENTRY.transfer.v1
INSTRUCTION.medication_order_schedule.v1
OBSERVATION.imaging_exam_image_series.v1
OBSERVATION.imaging_exam_report.v1
OBSERVATION.lab_test_single.v1

Discussion

Although all the clinical concepts retrieved from the CDR of Chinese hospital can be modelled using the openEHR archetype approach, several issues were encountered and discussed below.

Immaturity of archetype modification operations

openEHR specifies three operations for archetype modification: revision, specialization, and new version [15]. Other than these three operations, there is another operation with subtle differences. So far, the only way to add data items to existing archetypes is through specialization. Since each archetype represents a concept, this requires creating a new specialized archetype, which represents a new concept. There is no way to add these data items to existing archetypes. Since the archetype development is still an ongoing process, there are great demands for this requirements. We define a new operation for this situation named extension shown in Table 8.

Table 8 – Operations for archetype modification

Operation	Modification	Compatibility
Revision	Modify description part	Ensure backward compatibility
	Expand attributes, range of value sets, terminology	Data created by pre-revised archetype is compatible with the revised version
Specialization	Strengthen the constraints	Ensure the new specialized archetype must create data that conforms to the parent
	Redefine and add nodes	
	The range of value sets and semantics of nodes conform to the previous archetype	
New version	Change mandatory item to optional	Modifications are incompatible with the previous archetype
	Adjust value range or coded term set	
Extension	Add missing data items to existing archetypes	Compatible with the original archetype
	Use semantic version as the naming rule	

To take an example, if two data items of memo and report identifier need to be added in the archetype INSTRUCTION.request-imaging_exam.v1, it can be easily versioned as INSTRUCTION.request-imaging_exam.v1.1 through the "Extension" operation.

With the extension operation, archetypes can be clearly managed using the semantic versioning mechanism. Applying archetypes to local context will be much easier since there is no need to figure out a proper name each time to add data items through specialization.

The granularity of archetype

Differences in the granularity of archetypes between the CDR data requirements and CKM archetypes can cause problems of information representation in clinical practices. Take imaging examination sub-domain as an example. The concepts extracted from the CDR include Request, Request Item, Result, Report, DICOM Study, and Image shown in Figure 3a. Two coarse-grained archetypes (a) INSTRUCTION.request-imaging_exam.v1 which contains Request, Request Item and (b) OBSERVATION.imaging_exam.v1 which contains Result, Report, DICOM Study, Image are found in CKM and analyzed to extract corresponding concepts in Figure 3b. By comparing Figure 3a with Figure 3b, several important relationships between concepts are missing or improper, such as ③ one-to-many relationship between Report and Request Item, ⑤ one-to-one relationship between Request Item and DICOM Study, improper relationship cardinality between ④ Report and DICOM Study. The problem is that OBSERVATION.imaging_exam.v1 contains multiple concepts and the relationships between these concepts are greatly limited and not suitable for the CDR requirements. These differences result in structural modifications and reorganization of the archetypes which are demonstrated in Figure 3c. To represent ④, OBSERVATION.imaging_exam.v1 is split into two archetypes, (c) OBSERVATION.imaging_exam_report.v1 for Result and Report and (d) OBSERVATION.imaging_exam_image_series.v1 for DICOM Study and Image.

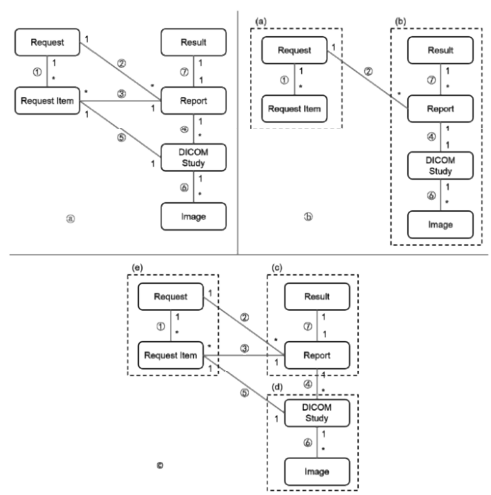


Figure 3 – Imaging examination concept relationships.

Bearing the principle of one archetype for one clinical concept in mind, with more and more clinical requirements appended, fine-grained archetypes is more flexible and easier to represent concepts and its relationships, and will improve the modelling capability of archetypes eventually.

Metadata-level versus data-level modelling

Problems are also encountered due to mismatches between metadata-level modelling and data-level modelling which happen in candidate archetypes and the CDR data requirements. For instance, all the results of the laboratory test have the similar structure, which consists of test item, test value and unit. In comparison to the clinical practice, for each laboratory test subject, such as full blood count and liver function, there is an archetype that contains certain data items of that subject. There are over 200 laboratory test subjects in the CDR, while only 19 of which have been defined with archetypes. Given the low scale of archetype coverage to clinical concepts, it is necessary to use an archetype that has the common structure for the general concept of laboratory test results.There is a OBSERVATION.lab_test.v1 archetype in CKM as the basis for all the laboratory test archetypes. We define a specialized archetype OBSERVATION.lab_test_single.v1 with data items test item, result, and result unit to represent the general strucutre of the laboratory test results.

Representing archetype relationships

Most of the archetypes in CKM use archetype slots to represent relationships between archetypes, but it has significant limitations that it can only express relation between certain types of archetypes. For example, archetype slot is not allowed between Entry based archetypes. The semantics of archetype slot is the major obstacle in this paper, and we choose links to express relationships between archetypes. For example, among imaging examination archetypes in Figure 3c, to relationship ④, a link node Image series is added to OBSERVATION.imaging_exam_report.v1 to refer to OBSERVATION.imaging_exam_image_series.v1. For ③, a node Report identifier is added under node activities to (e) INSTRUCTION.request-imaging_exam.v1.1. For ⑤, a link node Examination requested is added to OBSERVATION.imaging_exam_image_series.v1. Although link is a general method to represent relationships between archetyps, the usage of the link is not well explained in openEHR specifications and there are few examples. As relationship is an important aspect in information modelling, flexible relationship representation can greatly facilitate the application of the archetype approach to clinical practice.

Conclusion

This paper is the first research that builds a CDR and develops a clinical application with the openEHR methodology in clinical practice in China. A 6-step archetype modeling method was proposed, which refers to the MOH standards and exisiting database schemas for collecting required data items in CDR. It provides an important lesson for including experiments with the openEHR approach adoption in China, which also facilitates the openEHR adoption.

Although the feasibility of the openEHR methodology has been verified by the case study, some limitations of openEHR when implemented in China have also been identified, including immaturity of archetype modification operations, the granularity of archetype, metadata-level versus data-level modelling and representation of archetype relationships. The findings of the case study will facilitate the openEHR adoption in China.

Acknowledgments

This work was supported by the National High-tech R&D Program of China (2012AA02A601) and National Science and Technology Major Project of China (2013ZX03005012).

References

[1] Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. Journal of the American Medical Informatics Association 2010; 17(2): 131-5.
[2] Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. AMIA Annual Symposium Proceedings 2012; 2012: 189-198.
[3] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE–An integrated standards-based translational research informatics platform. AMIA Annual Symposium Proceedings 2009; 2009: 391-5.
[4] Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, Tao C, Parker CG, Haug PJ, Huff SM, Chute CG. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. Journal of biomedical informatics 2012; 45(4): 763-71.
[5] Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. AMIA summits on translational science proceedings 2010; 2010: 1-5.
[6] Beale T, Heard S. openEHR Architecture, Architecture Overview. Revision: 1.1. 2007.
[7] Beale T, Heard S, Kalra D, Lolyd D. The openEHR Reference Model, EHR Information Model. Revision: 5.1.1. 2008.
[8] Späth MB, Grimson J. Applying the archetype approach to the database of a biobank information management system. International journal of medical informatics 2011; 80(3): 205-26.
[9] Garde S, Hovenga E, Buck J, Knaup P. Expressing clinical data sets with openEHR archetypes: A solid basis for ubiquitous computing. International journal of medical informatics 2007; 76(3 Suppl): 334-41.
[10]Buck J, Garde S, Kohl CD., Knaup-Gregori P. Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach. International journal of medical informatics 2009; 78(8): 521-31.
[11]Zhang X, Yao Z. Modelling personal health record based on openEHR. Computer Applications and Software 2013; 30(5): 71-2.
[12]Li B. Discussion and research on the integration between SNOMED CT and opneER. Journal of medical informatics 2012; 33(1): 46-9.
[13]Yang Z, Liu D, Xu Y, Liang Y, Zhang H. The study of Clinical Information Model and Structured Clinical Data. Chinese journal of health informatics and management 2014; 11(1): 18-20.
[14]Xu W, Guan Z, Sun J, Wang Z, Geng Y. Development of an open metadata schema for prospective clinical research (openPCR) in China. Methods of information in medicine 2014; 53(1): 39-46.
[15]Beale T. Archetype Semantics. Revision: 0.5. 2007.

*Address for correspondence

Xudong Lu: E-mail address: lvxd@zju.edu.cn.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-212

Quality indicators from laboratory and radiology information systems

Matthieu Schuers[a,b], Mehr B. Joulakian[a], Nicolas Griffon[a,c], Joanne Pachéco[a], Carine Périgard[a], Eric Lepage[c,d], Ludivine Watbled[e], Philippe Massari[a], Stéfan J. Darmoni[a]

[a] *Department of Biomedical Informatics, Rouen University Hospital, Rouen Cedex, France*
[b] *Department of General Practice, University of Rouen, Rouen Cedex, France*
[c] *INSERM, U1142, LIMICS, Paris, France;*
Sorbonne Universités, UPMC Univ Paris 06 UMR_S 1142, LIMICS, Paris, France;
Univ. Paris 13, Sorbonne Paris Cité, LIMICS (UMR_S 1142), Villetaneuse, France.
[d] *Group for the development of shared Health Information Systems of Ile de France*
[e] *CIC-IT 1403, Lille University Hospital, France*

Abstract

Consequences of the computerization of laboratory and radiology information system (LIS and RIS) are not well documented. The aim of this study was to evaluate the impact of computerization of LIS and RIS of four hospitals on performance and quality of care. The study was divided into three phases. First, the subprocesses and information flows of LIS and RIS were described. Then, a literature review was performed in order to identify the indicators used to assess the impact of computerization. Finally, comparisons were made between 2 hospitals. Using the initial framework, each partner described its process mapping concerning LIS and RIS. The review identified a wide panel of indicators. Only 41 were useful to assess the impact of information systems. For each two by two comparison, lists of relevant indicators have been selected from the identified indicators and according to the process mapping comparison. Two by two comparisons have to be completed. Eventually, these indicators may be integrated in the quality process of hospital information systems.

Keywords:

Hospital information systems; process assessment (health care); quality indicators, health care

Introduction

Health information systems (HIS) have been widely studied in order to evaluate their effectiveness [1-4]. Benefits of HIS have been shown on the improvement of hospital productivity, coordination of care and quality of care [5]. The use of HIS also has a positive impact on the prevention of medication errors and the reduction of adverse effects [6]. However, most published studies are descriptive ones, and are performed in a monodisciplinary context. They do not provide the analysis of the consequences of the computerization implementation in a system involving multiple participants in the care process [7].

One important part of HIS is the computerization of laboratory and radiology services, from the prescription to the report of the exam results to the health professional. The computerization of laboratory and radiology services started many years ago [8], but still varies widely between health establishments [9].

The aim of this study was to: 1) describe the computerization of the subprocesses of a laboratory information system (LIS) and a radiology information system (RIS) of four hospitals; 2) identify indicators of the impact of computerization on performance and quality of care; 3) evaluate the impact of computerization on performance and quality of care using the pre-defined indicators.The same methodolody was applied to the three steps of this study, for LIS and RIS.

Methods

This study is part of the EVALSI project, led by a multidisciplinary consortium including four departments of biomedical informatics (Rouen, Paris, Lille and Nice), an engineer school (Mines ParisTech Graduate School) and a social sciences research centre (Research centre for the study and observation of life conditions).

The great variety of computerization between hospitals allowed us to build a global methodology based on comparisons between hospitals. This methodology was chosen in order to limit the risk of bias of before and after studies, following the advice of the scientific committee of the DHSO.

The study was divided into 3 phases. To date, phases 1 and 2 have been completed.

Phase 1: Process mapping

For each hospital, a detailed description of processes and information flow in the HIS of medical biological and imaging services was performed. For each process or sub process, the computerization level, the inter-step data flow, and the data that could be used for the measures of indicators were described. An initial framework was originally built from two hospitals (Lille and Rouen), and then used to obtain the final process mapping.

Phase 2: Identification of performance and quality of care indicators

In order to identify the indicators used to evaluate the subprocesses of medical biological and imaging systems, a literature review was performed. We used systematic search processes to identify all published studies concerning our topic. We searched the PubMed database, and selected articles in English and French. Detailed queries are available at: http://www.chu-rouen.fr/cismef/papers/Detailed_queries_used_for_the_literature_review.pdf. Inclusion criteria were studies performed in hospitals, concerning performance and quality indicators, used to evaluate the impact of HIS. The selection of articles was performed by two authors (JP and CP) and controlled by three others (MS, NG and SJD). We also systematically searched the reference lists of all the included studies and relevant reviews. Disagreements were resolved by consensus.

Phase 3: Evaluation of the impact of computerization on performance and quality of care

A list of performance and quality indicators was extracted from the literature review. Indicators that will be used for evaluation have been selected by five physicians (PM, MBJ, MS, NG, SJD), using 3 criteria: their applicability disregarding the level of computerization of subprocesses, their measurability disregarding the hospital and their relevance. Disagreements were resolved by consensus. Some data will be directly extracted from the HIS. Other data will be manually collected in laboratory and radiology services and in clinical wards (medicine and surgery). Comparisons will be made between 2 hospitals, process mapping of which showed a difference in the level of computerization in one subprocess.

Results

Phase 1: Process mapping

This phase has already been completed, and the results have been published. The result of the description of processes and subprocesses, using the previously developed framework, is a process mapping that allows for comparison between all the involved hospitals. Indeed, all participating hospitals described the same list of processes and subprocesses.

The following subprocesses have been described for the LIS: ordering, request filling, sample labelling, transmission of the request, sample delivery to the laboratory, registration of the request in the LIS, reconciliation of the request and the sample, technical validation, biological validation, report of the result, result awareness.

The following subprocesses have been described for the RIS: ordering, request filling, transmission of the request, appointment scheduling, organisation of the transport of the patient, registration of the patient in the RIS, production of the images, recording of the report, redaction of the report, transmission of the report, result awareness.

This process mapping allowed us to identify the steps where methods or levels of computerization were different between hospitals, and for which we may be able, using the indicators, to assess the impact of computerization (e.g presence of a ScanBack in hospital A vs. none in hospital B). The interested reader may refer to the publication for more information [10].

Phase 2: Identification of performance and quality of care indicators

Respectively, 446 and 986 articles were retrieved by the bibliographic queries for biology and radiology. Respectively, after reading the titles and abstracts (and the full texts in case of a doubt), 109 and 64 papers met our inclusion criteria (see figure 1).

Indicators for biological systems

The literature review included 109 articles. Examples of indicators reported in the literature are listed below by type.

- Delay indicators
  - Turnaround time and associated measures
  - Time between report of the result and reading of the result
  - Time between report of the result and consequential therapeutic change
  - Length of stay
- Quality indicators
  - Quality of the exam request filling
  - Proportion of samples that cannot be analyzed
  - Error rate in exam result reports
- Management indicators
  - Workload of the staff
  - Satisfaction of the staff, feeling of efficiency
  - Utilization rate of the machines
- Cost indicators

Indicators for imaging systems

This literature review included 64 articles. Examples of indicators reported in the literature are listed below by type.

- Delay indicators
  - Turnaround time and associated measures
  - Time between report of the result and reading of the result
  - Time between report of the result and consequential therapeutic change
  - Length of stay
  - Length of stay in medical imaging ward
- Quality indicators
  - Irradiation dose
  - False diagnostic rate
  - Rate of unread reports
  - Rate of lost reports
- Management indicators
  - Satisfaction of the staff, feeling of efficiency
  - Compliance rate with guidelines
  - Enhanced patient flow
- Cost indicators

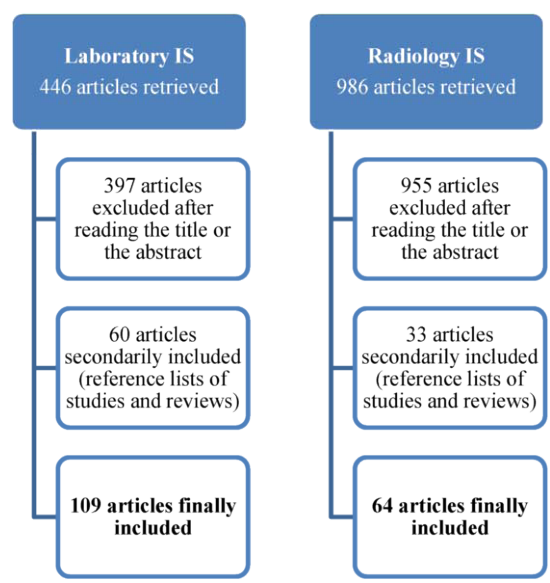


Figure 1 – Inclusion flow-chart

Phase 3: Evaluation of the impact of computerization on performance and quality of care

From the wild list of performance and quality indicators extracted from the literature review, two short lists of indicators for laboratory information and radiology

information systems were selected. Some activity indicators were also included in this selection. As a global indicator, turnaround time (TAT), which has been recognized in the literature as a key indicator [11], was impacted by all the subprocesses. Therefore, it has been decided to divide TAT into several sub-indicators. All the selected indicators are listed in tables 1 for biology and 2 for medical imaging.

Table 1 – Indicators selected for laboratory information systems (n=17)

Indicator type	Indicators
Activity indicators	Number of analyses over a given period
	Mean number of analyses per patient and per day
	Mean number of analyses per physician and per day
	Mean number of blood samples per hospital stay
	Mean number of blood samples per day of hospitalization
	Number of analyses per patient and per hospital stay
Performance indicators	Rate of cancelled analyses over a given period
	Rate of analyses rejected as non-compliant over a given period
	Time between sample collection and report of the result (TAT)
	Time between sample collection and arrival in the lab
	Time between arrival in the lab and start of analysis
	Time between start of analysis and technical validation
	Time between technical validation and biological validation
	Time between start of analysis and report of the result
	Time between the report of a drug dosage result and the consequential therapeutic change
	Time between the report of a hyperkalemia and the consequential therapeutic change
	Time between report of the result and reading of the result

Each indicator was precisely described, including all the data necessary to measure it. For each subprocess, indicators potentially impacted were identified. This allowed us to define the relevant indicators to perform each two by two comparison for LIS and RIS. Partial examples of the obtained matrixes are summarised in Tables 3 and 4, respectively.

In most cases, indicators were impacted by several subprocesses. The selection of indicators for each two by two comparison was automatically deducted from:

- the process differences between hospitals
- the potential impact of the computerization of subprocesses on the indicators.

The computerization of some subprocesses could be similar between two hospitals, for example, in Table 3, the ordering and sample labelling. The indicators only impacted by these subprocesses (eg: the rate of cancelled analyses over a given period) should be used as controls. Therefore, there should be no difference for these indicators between these two hospitals.

Table 2 – Indicators selected for imaging services information systems (n=24)

Indicator type	Indicators
Activity indicators	Number of exams over a given period
	Number of exams per patient and per stay relative to the length of stay
	Number of exams relative to the number of physicians
	Number of exams relative to the number of radiologists
	Number of exams relative to the number of available machines
Performance indicators	Number of redundant exams per hospital stay
	Compliance rate of exam requests with guidelines
	Proportion of exams requests modified by radiologists
	Proportion of images viewed by requesting physicians
	Proportion of reports read by requesting physicians
	Proportion of lost exams in patients files
	Absorbed radiation dose per patient
	Time between prescription and report of the result (TAT)
	Time between prescription and receipt of the request in the radiology service
	Time between receipt of the request and appointment schedulling
	Time between appointment schedulling and execution of the exam
	Timer between the scheduled time of the exam and the actual time of the exam
	Exam length
	Time between the end of the exam and the availability of the images for the requesting physician
	Time between the end of the exam and the availability of a first report for the requesting physician
	Time between the availability of the report and taking into account the results of the exam
	Time between the end of the exam and the availability of the final report for the requesting physician
	Patients waiting time in the radiology services
	Report writing time by the radiologist

In the case that one indicator was impacted by several subprocesses, and the difference in the computerization between two hospitals concerned only one of theses subprocesses, the above mentioned indicator should be considered as relevant to assess the impact of computerization of this subprocess. For example, in Table 4, the time between ordering and reception of the request in the medical imaging ward was potentially impacted by ordering, request filling and

Table 3 – Examples of computerization differences of several LIS subprocesses between Rouen and Paris hospitals, and their respective impacted indicators

Subprocesses	Rouen Hospital	Paris Hospital	Impacted indicators
Ordering	Computerized step (use of the LIS test catalogue)	Computerized step (use of the LIS test catalogue)	Number of analyses over a given period, mean number of analyses per patient and per day, number of cancelled analyses over a given period, number of analyses rejected as non-compliant over a given period, time between sample collection and report of the result, time between sample collection and arrival in the lab
Request filling	Paper application form	Computerized step	Time between sample collection and report of the result, time between sample collection and arrival in the lab
Sample labelling	Label identifying the patient	Label identifying the patient	Number of cancelled analyses over a given period, number of analyses rejected as non-compliant over a given period, time between sample collection and report of the result, time between sample collection and arrival in the lab
Transmission of the request	Physical transmission	Electronic transmission after validation	Time between sample collection and report of the result, time between sample collection and arrival in the lab

Table 4 – Examples of computerization differences of several RIS subprocesses between Rouen and Lille hospitals, and their respective impacted indicators

Sub processes	Rouen Hospital	Lille Hospital	Impacted indicators
Ordering	Computerized step (use of the RIS exams catalogue)	Paper prescription	Number of exams over a given period, number of exams per patient and per stay relative to the length of stay, number of exams relative to the number of physicians, number of exams relative to the number of radiologists, number of exams relative to the number of available machines, number of redundant exams per hospital stay, compliance rate of exam requests with guidelines, time between prescription and report of the result, time between prescription and receipt of the request in the radiology service
Request filling	Paper application form	Paper application form	Compliance rate of exam requests with guidelines, time between prescription and report of the result, time between prescription and receipt of the request in the radiology service
Transmission of the request	Physical transmission	Physical transmission	Compliance rate of exam requests with guidelines, time between prescription and report of the result, time between prescription and receipt of the request in the radiology service
Appointment scheduling	Partially computerized step	Manual registration	Number of exams over a given period, compliance rate of exam requests with guidelines, rate of exams requests modified by radiologists, time between prescription and report of the result, time between receipt of the request and appointment schedulling, time between appointment schedulling and execution of the exam

transmission of the request. Ordering is the only subprocess that differs between Rouen and Lille hospitals. Therefore, for the Rouen vs. Lille comparison, the time between ordering and reception of the request in the medical imaging ward was relevant to asses the impact of ordering computerization. Nevertheless, and still considering one indicator impacted by several subprocesses, if the difference in the computerization between two hospitals concerned more than one subprocess, the measures of the indicator should be hard to interpret, and could be useless.

Discussion

Using the initial framework, each partner described its process mapping concerning laboratory and imagery information systems. The literature review allowed us to identify a wide panel of indicators. Most of them were relevant to assess the quality of biological and medical imaging exam processes, but only few of them were useful to assess the impact of IS in our context. From the indicators identified and the process mapping, lists of relevant indicators have been defined to perform two by two comparisons.

TAT, and its subindicators, seemed to be well fitted to assess the impact of IS [11], and were considered quality indicators, especially in the biological domain.

Our study has several limitation.The standardized framework was quite basic. On the one hand, two quite different processes can be described in the same way. Therefore, it was necessary to have good knowledge of the process in order to avoid a false comparison. On the other hand, the framework facilitated the description of processes in hospitals as it did not require advanced skills for the description process. Moreover, the resulting process mapping was easy to compare as they shared the same format. The framework was adopted well by the four hospitals. Furthermore, only university hospitals were included in this study. This might be a possible concern for external validity. Nevertheless, their HIS profiles were quite different and many interesting comparisons could be made. A few subprocesses were similarly computerized. Obviously, this prevented the evaluation of the computerization of such processes, however, as indicators were impacted by multiple subprocesses, this strengthened the interpretation of indicators. The consortium validation of each step is a strength of this study: the validity of the overall process should be considered as strong.

These indicators will be integrated in the quality process of the Rouen LIS and RIS, and more generally in the Rouen computerized provider order entry and other involved applications of the HIS. Several other steps are already planned, in particular the extension of this quality process evaluation framework in other types of health facilities, eg: proprietary hospitals and nursing homes.

Acknowledgements

This study was supported by a grant from the French Ministry of Health (N°12-002-0002).

The authors thank Richard Medeiros – Medical Editor, Medical Editing International for editing the manuscript.

References

[1] President's Information Technology Advisory Committee. Revolutionizing Healthcare through information technology, 2004. http://www.nitrd.gov/pitac/reports/20040721_hit_report.pdf

[2] President's Council of Advisors on Science and Technology. Report to the president realizing the full potential of health information technology to improve healthcare for Americans: the path forward. December 2010. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

[3] Shekelle PG, Morton SC, Keeler EB. Costs and Benefits of Health Information Technology. Evidence Report/Technology Assessment, No. 132, Publication No. 06-E006. Agency for Healthcare Research and Quality, Rockville (MD), April 2006.

[4] Goldzweig Lubick C, Towfigh A, Maglione M, Shekelle PG. Costs and Benefits of Health Information Technology: New Trends from Literature. Health Aff (Millwood) 2009: 28(2): w282-93.

[5] Schoen C, Osborn R, Doty MM, Squires D, Peugh J, Applebaum S. A survey of primary care physicians in eleven countries, Perspectives on care, costs and experiences. Health Aff 2009: 28(6): 21171-83.

[6] Stroetman VN, Thierry JP, Stroetman KA, Dobrev A. eHealth for Safety - Impact of ICT on Patient Safety and Risk Management. European Commission, Information Society and Media. Office for Official Publications of the European Communities. Luxembourg, 2007.

[7] Huerta TR, Ford EW. Reflections on US health care information technology policy from a global perspective. Adv Health Care Manag 2012: 12: 111-20.

[8] Darmoni SJ, Massari P, Allaire P, Caffarel JL, Monconduit M, Baldenweck M, Hecketsweiler, P. REX: Extralaboratory Cycle Of Information Processing From Laboratories To Care Units. Proc. Annual International Conference of the IEEE Vol.14 Engineering in Medicine and Biology Society, Volume 3, Pages 1253-4, 1992.

[9] Atlas 2014 des SIH – état des lieux des systèmes d'informations hospitaliers (mai 2014) / DGOS – ATIH. http://www.sante.gouv.fr/IMG/pdf/Atlas_des_SIH_2014-2.pdf.

[10] Joulakian MB, Griffon N, Schuers M, Lepage E, Savoy-Collet C, Skalli S, Massari P, Darmoni SJ. Indicateurs en biologie et en imagerie au sein des systèmes d'information de santé. Journées Francophones d'Informatique Médicale 2014. http://www.chu-rouen.fr/tibs/wp-content/uploads/pdf/Joulakian2014.pdf

[11] Howanitz JH1, Howanitz PJ. Laboratory results. Timeliness as a quality attribute and strategy. Am J Clin Pathol. 2001 Sep;116(3):311-5.

Address for correspondence

Stéfan J. Darmoni

Department of Biomedical Informatics, Rouen University Hospital, 76031 Rouen Cedex, France

Stefan.darmoni@chu-rouen.fr

00 33 2 32 88 88 29

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-217

Building a Semantic Interoperability Framework for Care and Research in Fibromuscular Dysplasia

Marie-Christine Jaulent[a], Ariane Assélé-Kama[a], Sébastien Savard[b,d], Alessandra Giavarini[b], Emmanuel Touzé[c], Xavier Jeunemaître[b], Adrien Ugon[a], Pierre-François Plouin[b], Laurent Toubiana[a]

[a] INSERM, U1142, LIMICS, F-75006, Paris, France; Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France; Université Paris 13, Sorbonne Paris Cité, UMR_S 1142, LIMICS, F-93430, Villetaneuse, France
[b] Hypertension unit, Department of genetics and Centre for rare vascular diseases; Hopital Europeen G Pompidou (HEGP);Paris-Descartes University ; INSERM U970, Paris Cardiovascular Research Centre, 75015 Paris, France
[c] Université Caen Basse Normandie, CHU Côte-de-Nacre, service de neurologie et unité neurovasculaire, 14000 Caen, France; INSERM U919, GIP CYCERON, 14074 Caen, France.
[d] Hôtel-Dieu de Québec, Service de néphrologie, Quebec, Canada

Abstract

Identifying patients with Fibromuscular Dysplasia (FMD) at the international level will have considerable value for understanding the epidemiology, clinical manifestations and susceptible genes in this arterial disease, but also for identifying eligible patients in clinical trials or cohorts. We present a two-step methodology to create a general semantic interoperability framework allowing access and comparison of distributed data over various nations, languages, formats and databases.
Methods: The first step is to develop a pivot multidimensional model based on a core dataset to harmonize existing heterogeneous data sources. The second step is to align the model to additional data, semantically related to FMD and collected currently in various registries. We present the results of the first step that has been fully completed with the validation and implementation of the model in a dedicated information system (SIR-FMD). We discuss the current achievements for step 2 and the extensibility of the methodology in the context of other rare diseases.

Keywords:

Information Systems; Health Information Exchange; Semantics; Vocabulary; Fibromuscular Dysplasia.

Introduction

Fibromuscular dysplasia (FMD) is a heterogeneous group of idiopathic, non-atherosclerotic, relatively rare vascular diseases, leading to the narrowing of medium-sized arteries, mostly the renal and internal carotid arteries [1], [2]. The etiology and pathogenesis of this disease are unknown and there is no biochemical test for the condition. The diagnosis and classification of FMD relies today on angiographic findings. These data are collected in heterogeneous databases at different points of care. Moreover, additional data about patients are often associated with these findings but not routinely. For instance information about "Anisuporia / non-reactive pupils" is collected in the American registry (FMDSA[1]) [3], but not systematically collected in other databases. Beside this, the large diversity of cultures, laws, regulations and operational implementation regarding personal health data processes (access, gathering, sharing, etc.) across countries induces variation in the nature of the data that are collected. For instance the "race" information will be systematically present in USA registries while in European ones, the information will be "geographical origin of the parents".

Frameworks accessing and comparing distributed and heterogeneous data with consistent semantics are needed to enable cooperative research and progress in the comprehension of the disease. Our main challenge is to develop such a framework that includes an e-registry collecting standardized FMD data across Europe and that gives access and reuse of existing local FMD registries in other regions. E-registries have been developed by the authors in other domains [4][5].

We have designed a generic two-step methodology to set this framework in the context of FMD based on current Information Technology paradigms.

Nowadays, semantic interoperability is a broadly used paradigm to approach the problem of sharing data from heterogeneous data sources [6]. Many semantic interoperability platforms have been developed in various projects to apply this paradigm for data analysis by querying heterogeneous and distributed data sources. Examples are given by the EHR4CR European project [7] and the DebugIT European project [8]. The rationale of such systems is to define the intended meaning (semantics) of the data to ensure coherent interpretation by humans and processing tools, even when data are not coded the same way [9]. Some existing frameworks are based on an ontology to unify structural models and terminologies together with relevant mapping sets. This approach has been tested in the context of the EU Framework Program 7 TRANSFoRm project [10].

In the following, we describe the two-step method to develop a framework that comprises first, a European e-registry based on a pivot multidimensional model to merge existing local FMD databases in France and second, mapping resources to make the registry usable at the international level by sharing data semantics with the USA FMD registry. We present the resulting multidimensional model and the current status of the project.

Finally, we discuss some issues raised by this two-step methodology and its generalisability.

[1] http://www.fmdsa.org/research_network/fmd_registry accessed on December 14, 2014

Materials and Methods

Material

Since 2006, the French authorities have implemented a network of 131 rare disease centers of expertise in the country. The reference center for FMD is located in the "Georges Pompidou European Hospital" (HEGP), Paris, France. In collaboration with the Hypertension Unit of HEGP, the reference center started the collection of FMD data in 2009. Clinical and biological data are collected at the first visit. For patients referred to the center after FMD has been diagnosed elsewhere, clinical and biochemical data are collected at the first visit in the center if it occurred within 1 year of the diagnosis of FMD and if there had been no renal artery intervention during this period.

A French network of nephrologists, neurologists, radiologists and specialists in hypertension is responsible to document phenotypic and genetic traits of the disease and the progression of FMD lesions in patients with renal and/or cervical artery FMD. Information regarding >500 patients with FMD is currently scattered in several non-standardized, redundant databases (Figure 1).

1. The FMD local database of the HEGP reference center in Paris contains about 400 cases described each by *300 items*.
2. A database validated by the HEGP reference center and dedicated to activity reporting at the national level includes about *50 items* mostly redundant with the local database.
3. The ARCADIA/PROFILE program[2] was completed in October 2014. A specific Case Report Form was defined in this programme. The resulting database contains 500 patients with renal artery or cervical artery FMD from 12 different clinical sites (including HEGP) and described each by *200 items*[3].

In 2007, the Fibromuscular Dysplasia Society of America (FMDSA) decided to begin a registry to better understand the disease and its treatment. A template has been defined to collect *318 items*. Clinical data include elements such as date of diagnosis, types of tests conducted and results of these tests, past medical history, family history, subsequent clinical events and any clinical outcomes [2].

Besides this, several smaller studies exist in Europe to answer specific questions concerning the patients and reported in scientific papers. For these studies, specific databases, most of the time Excel files, are locally built for data management and analysis [1]. Such databases are redundant with ARCADIA/PROFILE but may include interesting additional items.

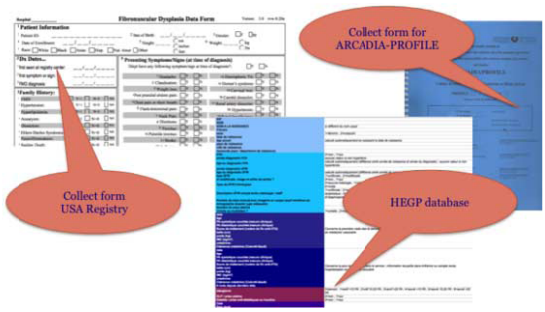


Figure 1 – Heterogeneous databases to collect FMD patients in France and USA

The HEGP FMD center wants to propose a national French e-registry based on the Case Report Form used in the ARCADIA/PROFILE cohort. It also aims at extending this registry to European countries and to making it compatible, as far as possible, with the USA registry.

Methods

Ideally, data from multiple sources are converted to standardized formats using well-characterized data vocabularies. Such approach currently requires huge efforts from experts to reach a consensus about the semantics of the domain and the standard vocabulary. It is for instance the case with the National Database for Autism Research[4] who reaches consensus about the semantics of the data.

In the first step of our methodology, a multidimensional model is set for the definition of a standard European FMD e-registry through a close collaboration between clinical experts and ontologists. This model is based on the identification of a set of core data elements for which a consensus is reachable with reasonable efforts. The second step consists of mapping all the items of the local database to the USA model. Figure 2 draws the two-step methodology.

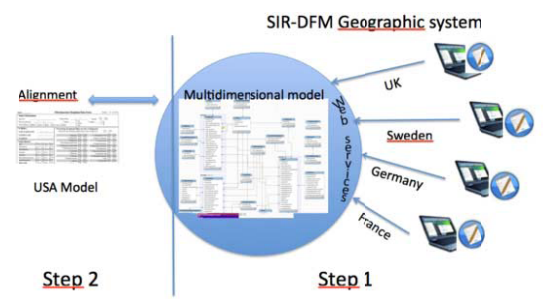


Figure 2 – The two-step methodology to set the framework

A bottom-up approach to build a multidimensional model

A data set includes data elements (DE) corresponding to some specifications which stipulate the sequence of inclusion of the DEs, whether they are mandatory, what verification rules should be employed, and the scope of the collection. A core data element (CDE) is a DE that is used in various data sets and recognized by the experts as a standard information. A specific DE (SDE) is a DE specifically defined for a given purpose (eg, the "race" is a SDE used as a public health indicator in the US) [11].

The US Office of Rare Diseases Research (ORDR) developed a set of CDEs at a national level as a tool for global collection of rare disease patient data [12]. In France, a minimum dataset

[2] ARCADIA, Assessment of Renal and Cervical Artery DysplasIA
[3] PROFILE, PROgression of FIbromuscular Lesions; Programme Hospitalier de Recherche Clinique, French Ministry of Health 2009-2014

[4] http://ndar.nih.gov. Accessed January 28, 2014.

for rare diseases has been set to respond to public health and epidemiological queries across all rare diseases [13]. These initiatives demonstrate the value of building CDE data sets. In [13], the authors argue the necessity to define core minimum data sets for each group of rare diseases in order to continue the efforts to make data interoperable for research.

Excel files and paper-based clinical Research Forms were identified and analysed by both FMD experts and ontologists to reach a consensus on a conceptual model for FMD descriptions. The target FMD model is a fixed set of CDEs. In terms of standardisation, the coding of the medical domain values can be adapted to a country's existing practices. For example, the rare disease diagnosis can be coded in Orphanet[5], OMIM, or SNOMED CT depending on the context and the required granularity of the coding. In France, diagnostic coding relies on Orphanet codes and is done by clinicians. Academic portals such as UMLS[6] or CISMeF[7] provide alignments between coding systems.

We built a set of fake cases, each case corresponding to a virtual patient, and asked a physician to describe the case using the model in order to confirm its accuracy. An information system implementing the model has been developed and is currently routinely used to feed the e-registry (Figure 2, step 1).

Using reference terminologies to align SDEs

One limit of using CDE come from the fact that, very often, data are collected at a local level for specific studies and may include specific data, semantically related to FMD but not explicitly expressed in the multidimensional model. For instance, the item "Summary_Smoking" is collected in [1] to answer a given query, it is semantically related to the item "number-of-cigarettes_daily" of the multidimensional domain but some alignment is needed. In this case, semantic interoperability can be achieved through data integration guided by an ontology [9] and in particular the approach of the type "global as view" in which an overall ontology is used as a source of mediation. However, and although there are starting efforts in the field, a domain ontology does not exist yet for FMD. In this situation, reference terminologies developed by Standard Development Organisations may be used to explicit the semantics of the data. The important initiative CTS2[8] is a specification for representing, accessing and disseminating terminological content. Moreover, many services exist to provide correspondences between these reference ressources [14].

Such an approach is used in the second step of our methodology to define mappings between reference terminologies and local interface terminologies used in local registries.

Results

Step 1 : e-registry

A standardized set of Core Data elements

The design of the methodology resulted in the following steps: (i) collecting a flat list of items as the union of the data elements and data values available from the material; (ii) performing a systematic review of the items by a group of experts; (iii) setting the different tables and value sets corresponding to the items and (iv) standardization by the eHealth team using Standard Development Organisations resources to support semantic interoperability and facilitate data re-use.

This step allows us to update the FMD nomenclature (e.g. unifocal vs. multifocal.) and to standardize the definition of FMD and FMD subtypes in order to harmonize clinical practices and enable cooperative research for this pathology.

Several face-to-face meetings were conducted with experts to reach a compromise and obtain the set of CDEs. The USA registry items were considered during this process in order to cover as many items as possible that were collected from the material. The CDE set is composed of 194 items.

The standardization of the data elements consists of:

1. Identifying concepts in reference terminologies that correspond to the items. This allows us to standardize the value sets associated to the items.
2. Choosing the prefered terms for the items and the elements in the value sets.

For example, the item "Artères" is linked to the concept "Set of Arteries" in the FMA reference ontology[9] and all elements in the value set of "Artères", such as "Coronaire Droite" are also associated to concepts in FMA (see Figure 3).

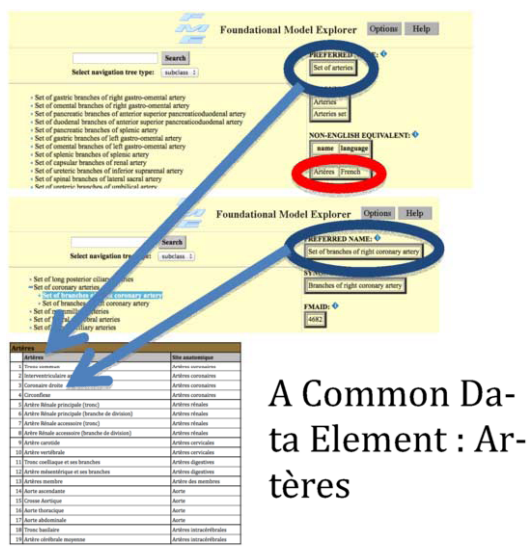


Figure 3 - Example of a common Data Element "Artères" resulting of an experts consensus and standardized with FMA concepts

The resulting multidimensional model is composed of three fact tables corresponding respectively to "Diagnosis", "Past Medical History" and "Current Situation". It is composed of 20 dimensions which are: Patients, Geographical Origin; Link; Symptoms; Signs; Vascular Events, Vascular Event Origin; Vascular Risk Factors; Type of FMD injury; Type of FMD histology; Surgery/Intervention; Type of intervention; Type of exam; Imagery Interpretation; Anatomical site, Arteries; Localisation; Drug Family; Drugs; Physical Exam; Biochimie, Familial History. Two groups of items have been designed in addition (exams; therapies) to take into account temporal and iterative information.

The Figure 4 highlights the table Diagnosis and the relations with the other tables.

[5] http://www.orpha.net/consor/cgi-bin/index.php
[6] http://www.nlm.nih.gov/research/umls/
[7] http://www.chu-rouen.fr/cismef/
[8] http://informatics.mayo.edu/cts2/index.php/Main_Page
[9] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

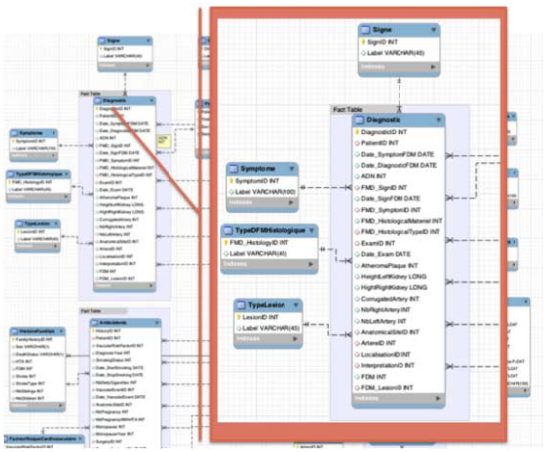


Figure 4 - Sample of the multidimensional model: the table Diagnosis

Development of the e-registry

The SIR-FMD information system was developed in 2013 with the "MetaSurv" generator described elsewhere [15]. It implemented the multidimensional model described above. A secure platform has been made available as an evaluation platform since the beginning of November 2013. On the user side, SIR-DFM relies on existing local Internet networking facilities. Via a web browser, the user connects to the interface, which is connected to the databases. This application is organized into multiple specifications that provide access to intelligent electronic forms. These electronics forms are scalable, i.e., it is possible at any time to add new items that can be used for any other patients.

SIR-DFM fulfils several requirements: scalability, portability, reliability, accessibility and cost-effectiveness oriented toward non-proprietary software. After the validation of the model, medical records of 471 patients from the databases presented in the Material section of this paper were included and are accessible through a secure user account. Users are organized into a collaborative group, and can access patient groups. The system has been presented recently in the context of the international Fibromuscular Dysplasia Research Network symposium in Cleveland, Ohio USA [16]. At first, this platform is only accessible by members of the French network, but by design, the extension of this registry to European countries is possible and depends only on the adoption of the multidimensional model at the European level.

Step 2 : Alignment process for SDEs

The proposed e-registry model includes a set of Core Data Elements with unambiguous definitions. It would be complex, due to legal and regulatory differences between Europe and USA, to settle on a single international e-registry. Some items present in the USA forms are not present in the FMD e-registry, like for instance, race or religion. One important aspect is to be able to query both data sources (to include patients in a cohort for instance) and then to establish meaningful alignments between items of the different databases that are not in the CDE dataset.

Automated schema alignment approaches have been released to avoid spending time on manual alignment to detect similarity between data elements[10]. The way to classify these alignment approaches may differ in the literature [17][18]. In our current context, the alignment cannot be automatic due to the lack of an Ontology in the FMD domain. We have analysed the potential alignments on a case by case basis. Some propositions, like the "race" item in the USA and the "geographical origin" item in France,are identified as semantically linked. At first, we manually fixed ad-hoc the similarity degree between the identified correspondences.

The mapping process has three levels of correspondences: 1) the table section (21 sections in France and 10 sections with a lot of sub-sections in USA); 2) the items level with almost double items in the USA registry; 3) the value sets.

Figure 5 shows an example of the mapping effort at the level of items and value sets. We followed the methodology described in [18]. The work is in progress and accounts for the building of the semantics of the domain.

HEGP			
Table	**num**	**items**	***Description***
Patient	1	NIP	
	2	NOM	
	3	NOM A LA	*si différent du nom usuel*
	4	Prénom	
	5	sexe	*1=féminin ; 2=masculin*
	6	date de naissance	
	7	âge actuel	
	8	pays de naissance	
	9	ville de naissance	
	10	hexacode pays / département de naissance	
USA			
Table	**num**	**items**	***Description***
Patient	1	Patient ID	
	3	Gender	*F or M*
	2	Date of Birth	*dd/mm/yyy*
	4	Date of Enrollment	*dd/mm/yyy*
	5	Height	*units: cm, inches, feet*
	6	Weight	*units:kg, lbs*
	7	Race	*white, black, asian, hisp, nat. Amer, other*

HEGP			**USA**		
Type d'examen	1	Angio-Scanner	*Exam Type*	1	Angiogram
	2	Angio-IRM		2	CCA/ICA Plaque
	3	Arthérographie		3	Ultrasond
	4	DIVA		4	Transcranial Doppler
	5	Echographie			

Figure 5 - Sample of correspondances set between the e-registry and the USA registry

Discussion and Conclusion

Fibromuscular dysplasia is a heterogeneous group of rare vascular diseases, leading to the narrowing of medium-sized arteries, mostly the renal and internal carotid arteries. Patients data are today collected in heterogeneous databases at different points of care. In order to enable cooperative research and progress in the comprehension of the disease, our challenge is to develop a framework to access and compare such distributed and heterogeneous data with consistent semantics. We propose a generic two-step methodology to

[10] ISO/EIC 11179 s Information technology - Metadata registries (MDR) - Part 3:Registry metamodel and basic attributes, Third edition 2013-02-15

develop this framework. We have fully completed the first step of the methodology. A multidimensional model was developed from existing material and working sessions with experts and ontologists. An information system was implemented, SIR-FMD, based on this dimensional model to set an e-registry. Currently, all pre-existing patients from the ARCADIA/PROFILE program are included in the e-registry. A prospect of this work is to extend this e-registry to European countries as they will increase their use of the SIR-FMD platform [19]. One important further step will be to support the integration of EHR data in the SIR-FMD platform.

The second step is in progress and needs a large group of experts. A collaboration between the French team and the FMDSA team started recently with two objectives: 1) validate the correspondences between SDEs proposed by the French team and 2) agree on semantics. More work has to be done at the level of the SIR-FMD platform to integrate all needed alignments for performing new research queries across the European e-registry and the FMDSA registry.

Finally, the methodology can be easily extended in other contexts. For instance, we are currently developing an international cohort for the ehlers danlos syndrome, a different rare disease, re-using the methodology and the "MetaSurv" generator.

Acknowledgments

We thank *Jeffrey Olin*, Professor of Medicine at the Icahn School of Medicine at Mount Sinai Hospital and Director of Vascular Medicine and the Vascular Diagnostic Laboratory in the Zena, New York, as well as *Pam Mace*, President of the Fibromuscular Dysplasia Society of America for giving us access to the USA FMDA questionnaire.

References

[1] Savard S, Steichen O, Azarine A, Azizi M, Jeunemaitre X, Plouin PF. Association between 2 angiographic subtypes of renal artery fibromuscular dysplasia and clinical characteristics. Circulation. 2012 Dec 18;126(25):3062-9.

[2] Olin JW, Froehlich J, Gu X, Bacharach JM, Eagle K, Gray BH, Jaff MR, Kim ES, Mace P, Matsumoto AH, McBane RD, Kline-Rogers E, White CJ, Gornik HL. The united states registry for fibromuscular dysplasia: Results in the first 447 patients. Circulation. 2012;125:3182-3190

[3] Poloskey SL, Olin JW, Mace P, Gornik HL. FMD Patient Page. Circulation 2012;125:e636-e639.

[4] Gimenez-Roqueplo AP, Lehnert H, Mannelli M, Neumann H, Opocher G, Maher ER, Plouin PF; European Network for the Study of Adrenal Tumours (ENS@T) Pheochromocytoma Working Group. Phaeochromocytoma, new genes and screening strategies. Clin Endocrinol (Oxf). 2006 Dec;65(6):699-705.

[5] Landais P, Messiaen C, Rath A, Le Mignot L, Dufour E, Ben Said M, Jais JP, Toubiana L, Baujat G, Bourdon-Lanoy E, Gérard-Blanluet M, Bodemer C, Salomon R, Aymé S, Le Merrer M, Verloes A; CEMARA task force. CEMARA an information system for rare diseases. Stud Health Technol Inform. 2010;160(Pt 1):481-5.

[6] Martínez-Costa C, Kalra D, Schulz S. Improving EHR semantic interoperability: future vision and challenges. Stud Health Technol Inform. 2014;205:589-93.

[7] Moor GD, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, Karakoyun T, Ohmann C, Lastic PY, Ammour N, Kush R, Dupont D, Cuggia M, Daniel C, Thienpont G, Coorevits P. Using Electronic Health Records for Clinical Research: the Case of the EHR4CR Project. J Biomed Inform. 2014 Oct [Epub ahead of print]

[8] Lovis C, Colaert D, Stroetmann VN. DebugIT for patient safety - improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. Stud Health Technol Inform. 2008;136:641-6

[9] Assélé Kama A, Choquet R, Mels G, Daniel C, Charlet J, Jaulent MC. An ontological approach for the exploitation of clinical data. Stud Health Technol Inform. 2013;192:142-6.

[10] Ethier JF, Dameron O, Curcin V, McGilchrist M, Verheij R,Arvanitis T,Taweel A, Delaney B, Burgun A. A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm. J Am Med Inform Assoc 2013;20:986–994.

[11] Winget MD, Baron JA, Spitz MR, et al. Development of common data elements: the experience of and recommendations from the early detection research network. 2003;5056:41–8.

[12] Forrest CB, Bartek RJ, Rubinstein Y, et al. The case for a global rare-diseases registry. Lancet 2011;377:1057–9.

[13] Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc. 2014

[14] Grosjean J, Merabti T, Soualmia LF, Letord C, Charlet J, Robinson PN, Darmoni SJ. Integrating the human phenotype ontology into HeTOP terminology-ontology server. Stud Health Technol Inform. 2013;192:961.

[15] Toubiana L, Moreau S, Bonnard G. MetaSurv: Web-platform generator for the monitoring of health's indicators and interactive geographical information system. Stud Health Technol Inform. 2005;116:989-93.

[16] P.-F.Plouin, L. Toubiana, M.-C. Jaulent, X. Jeunemaitre French FMD registry: organizational structure, data elements, summary of findings. International Fibromuscular Dysplasia (FMD) Research Network symposium May15th and 16th 2014 Cleveland, Ohio USA

[17] Rahm E, Bernstein PA. A survey of approaches to automatic schema matching, The VLDB Journal 10, 334–350, 2001.

[18] Maaroufi M, Choquet R, Landais P, Jaulent MC. Formalizing mappings to optimize automated schema alignment: application to rare diseases. Stud Health Technol Inform. 2014;205:283-7.

[19] Persu A, Giavarini A, Touzé E, Januszewicz A, Sapoval M, Azizi M, Barral X, Jeunemaitre X, Morganti A, Plouin PF, de Leeuw P; ESH Working Group Hypertension and the Kidney. European consensus on the diagnosis and management of fibromuscular dysplasia. J Hypertens. 2014 Jul;32(7):1367-78

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-222

Micro- and Macrointegration Profiles for Medical Devices and Medical IT Systems

Raluca Pahontu[a], Angela Merzweiler[b], Gerd Schneider[a], Björn Bergh[a]

[a] *Center for Information Technology and Medical Engineering, Heidelberg University Hospital, Heidelberg, Germany*
[b] *Integrating the Healthcare Enterprise, German Deployment Committee, Berlin, Germany*

Abstract

Collecting, saving, and providing patient data are essential processes of documentation in a hospital. Many IT systems have evolved to provide solutions in this area. The automatic transfer of medical device data to these information systems is a new challenge for hospital IT systems. Some vendors are focused on the integration of medical IT systems and medical devices. They provide great solutions with magnificent features. Nevertheless, those integration solutions are proprietary and isolated, limiting the operator's selection of his medical devices and medical IT systems. Standardizing communication processes within the operating room and between medical devices and medical IT systems brings benefits for both patient and hospital staff. This work identifies and proposes micro- and macrointegration profiles as a basis for new IHE Integration Profiles for both medical IT systems and medical devices of the operating room.

Keywords:

Standardization; IHE; healthcare; medical devices; interoperability; Hospital IT Systems.

Introduction

The ongoing spread of technology in the healthcare domain increases the necessity of connecting medical devices and clinical information systems. In healthcare, interoperability is the ability of different information technology systems and software applications to exchange data and to interpret these data in the same way as defined in the HIMMS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations [1]. Semantic interoperability between medical devices and clinical information systems will permit mixed-vendor data transfer and comprehensive secure data acquisition, and it will enable innovations to improve patient safety, treatment, and workflow efficiency, reducing medical errors and healthcare costs to the benefit of patients [2].

The main goals of the OR.NET [3] project–funded by the German Federal Ministry of Education and Research (BMBF)–are the development of semantically interoperable, standardized, certifiable, secure, multi-vendor networking solutions for medical devices and clinical IT systems for the operating room and clinic. In order to achieve technical, syntactic, and semantic interoperability, OR.NET supports the Integrating the Healthcare Enterprise (IHE) initiative. IHE is an initiative of healthcare professionals and system vendors focused on improving the way computer systems share information [4]. IHE supports established standards such as Digital Imaging and Communications in Medicine (DICOM) [5] and Health Level Seven (HL7) [6] to address specific clinical needs, focusing on real-life integration problems. Providing specifications on implementing existing standards, tools, and services for interoperability, IHE establishes a guideline for a standardized communication in several clinical domains like IT infrastructure, radiology, patient care devices, and much more. It addresses aspects like security, patient privacy, and safety and shows how required data can be exchanged across different actors. Even if existing standards address some communication needs, they cannot close all interoperability gaps. This paper identifies gaps of the standards and promotes micro- and macrointegration profiles modelling communication needs between devices of operating rooms and clinical information systems.

Materials and Methods

The development of new IHE Integration Profiles usually consists of four steps [7]:

1. Clinical and technical experts define critical use cases for information sharing.
2. Technical experts create detailed specifications for communication among systems to address these use cases, selecting and optimizing established standards.
3. Industry implements these specifications called IHE Profiles in hospital information systems (HIT).
4. IHE tests vendors' systems at carefully planned and supervised events called Connectathons.

During the OR.NET project, various methods were used to identify communication needs in the operating room and clinic.

As a first step, we collected various clinical and technical use cases describing clinical and technical workflows between the operating room and medical IT systems in a standardized manner. Each use case should be characterized by the following information: author, involved actors, triggering events, preconditions, postconditons, invariants, standard workflow, alternative workflow, and temporal requirements of the arrival of transferred data. Temporal requirements range from closed-loop control of interconnected medical devices requiring real-time communication to archiving of patient images which is not critical in terms of time. Furthermore, each use case was linked to corresponding exchanged data in a communication matrix. All partners of the project (physicians, IT system vendors, medical device vendors, etc.) were involved in this process. We collected about 150 use cases.

As a further step, we developed a communication matrix template to investigate which devices or IT systems should communicate with each other. Medical devices and medical IT systems were classified in these categories:

- Anesthesia Devices: Anesthesia Machine, Ventilator, Infusion Pumps, Syringe Pumps, Heat Preservation Mats, Patient Monitoring, Neurophysiological Monitoring, Anesthesia Workstation
- Imaging Devices: Ultrasound Device, X-Ray, CT Scanners, MRI Scanners, Surgical Microscope Camera Ceiling
- Navigation: Navigation System, Tracking Camera
- Other Tools: Irrigators, Aspirators, Foot Pedal, Surgical Table
- Building Utilities: Ambient Light, Displays, Air Cooling
- IT-Systems: Clinical Information System (CIS), Documentation System, Patient Data Management System (PDMS), Picture Archiving and Communication System (PACS)

These categories are represented by the x and y axis of our communication matrix.

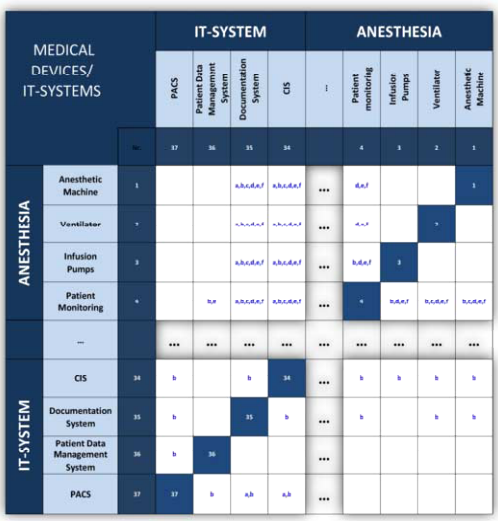


Figure 1 – Fragment of communication matrix

Furthermore, we classified possible exchange data in order to investigate which kind of data should be exchanged by these devices and systems. Therefore, we classified data into:

- Patient-related data
  - Streaming Data: Video (e.g. Endoscope-, Laparoscope videos); Measured values (e.g. heart rate, blood pressure, patient temperature); Navigation data
  - Non-Streaming Data: Patient administration data; Images (e.g. model data, segmented data); Discrete measured values (e.g. heart rate, blood pressure, patient temperature, medical report, laboratory findings, surgical planning data)
  - Alert data: like ventilation pressure is too high
- Non-patient-related data
  - Streaming Data: Room temperature; Medical device's status information; Navigation data (e.g. device position in the room); Device configuration parameters
  - Non-Streaming Data: Discrete measured values (e.g. 30% brightness, table's tilt); Medical device's status information; Device configuration parameters
  - Alert data: like technical alarms

Different project partners, such as research organizations, IT system developers, medical device vendors, and physicians, filled the matrix template with specific patient-related and non-patient-related data. As a result, we aggregated 16 filled matrices in one communication matrix. Figure 1 shows a fragment of the aggregated matrix.

Based upon the 150 concrete use cases and the communication matrix we derived four abstract microintegration profiles (MIP) describing four main communication types between clinical IT systems and medical devices or between medical devices. Each microintegration profile consists of the involved actors and a name for the transaction. The four microintegration profiles can be composed to one abstract macrointegration profile (MAP).

As a fourth step, we performed a literature research on existing IHE profiles to map identified communication needs to existing IHE transactions.

Results

The following results were obtained from the communication matrix:

- We identified the medical devices and IT systems that need to communicate with each other. In IHE terms, those also represent the actors of the microintegration profiles.
- We identified the kinds of data that are exchanged by those devices or IT systems
- We clustered all identified communication needs [8], resulting in four main MIPs and an abstract MAP. These profiles are described separately below.

MIP "Communicate Worklist"

The first MIP addresses the transfer of universal worklists to medical devices within an operating room. Thus, users are able to access and manipulate work items (e.g. orders, tasks, requests) originating from various IT systems, such as an operation planning system. The universal worklist can be accessed via a user interface that can be customized so that users can get only the most important information about the execution of the tasks displayed. For example, physicians can query all of the day's upcoming surgery orders and patient names. This worklist could then be used to select a patient for documentation. In order to communicate the worklist we need a worklist provider and a worklist consumer. In terms of IHE, those will represent the actors of the transaction called Communicate Worklist as shown in Figure 2.

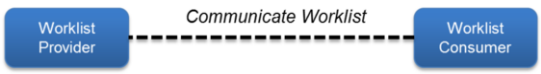


Figure 2 – MIP Communicate Worklist

MIP "Communicate Patient Demographics"

Since not every system provides the functionality of communicating a worklist, we propose a fallback mechanism if the universal worklist is not available. Therefore, the patient's demographic data have to be digitally transmitted from the clinical information system to the operating room as shown in Figure 3. In this case, the two actors will be a patient demographics provider and a patient demographics consumer. "Communicate patient demographics" is the needed transaction.

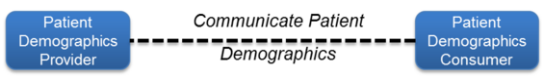


Figure 3 – MIP Communicate Patient Demographics

In our scenario, the microintegration profiles Communicate Worklist and Communicate Patient Demographics mainly refer to a communication between the operating room and medical IT systems.

MIP "Communicate Results"

During surgery medical devices deliver both patient-related and non-patient-related data. Those result data are relevant for the involved devices as well as for clinical subsystems like an operation planning system or a clinical information system. This communication path is displayed in Figure 4.

Figure 4 – MIP Communicate Results

For the MIP Communicate Results we first classified the results that should be communicated in:

1. Multimedia results like images, videos etc.
2. Alphanumeric results like diagnostic data, examination parameters etc.

Depending on the storage target, results should be available in raw, filtered, and aggregated form. Medical devices and IT systems may implement both actors: a result provider and a result consumer. For example, a PACS provides multimedia objects like patient images and saves images from imaging devices as a result consumer. A patient monitoring system may provide vital data and consume patient demographics from a patient demographics consumer actor.

MIP "Communicate Technical Operation and Control Parameters" (TOCs)

Some clinical use cases also address communication in the operating room between medical devices like the communication of technical operation and control parameters. TOCs can, for example, represent the current status of medical devices like light on/off or control parameters like light endoscope brightness increased by 30%. The communication of TOCs between an ultrasound dissector (USD) and a surgical microscope is one of OR.NET's use cases. In this scenario the physician can view and modify USD parameters via the user interface of a surgical microscope. Here we distinguish between read operations like data visualization, read/write control, and closed-loop control. Closed-loop refers to the feedback interconnection between two medical devices, meaning that a medical device action is dependent on the deviation between the measured and desired performance. Read, read/write control, and closed-loop control have various temporal requirements that imply different communication transport media. While for visualization data and read/write control Ethernet network technology may satisfy the corresponding temporal requirements, a real-time network technology is mandatory for closed-loop control data so that patient safety can be guaranteed every time. Therefore, this MIP has two specifications depending on the required network technology: Communicate TOCs in non-real-time and Communicate TOCs in real-time. Figure 5 shows the abstract MIP Communicate TOCs.

Figure 5 – MIP Communicate Technical operation and control parameters

Figure 6 shows one macrointegration profile (MAP) as a result of the combination of the four MIPs.

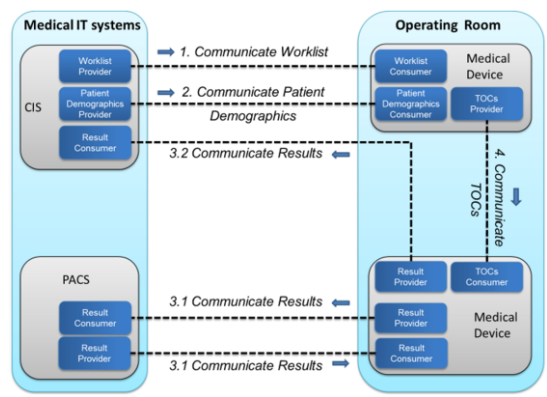


Figure 6 – Macrointegration profile

This abstract MAP includes the following microintegration profiles.

- MIP Communicate Worklist
  - Alternative: MIP Communicate Patient Demographics
- MIP Communicate Results
  - Multimedia results
  - Alphanumeric results
- Optional: MIP Communicate TOCs in non real-time
- Optional: MIP Communicate TOCs in real-time

Depending on the clinical scenario or the imminent surgery, the microintegration profiles can be variously combined, respecting the process data flow like starting with patient identity data, communicating results to medical devices and to medical IT systems. As an example, one specific MAP based on the communication between CIS, PACS, the Möller-Wedel GmbH & Co KG surgical microscope, and Söring GmbH

USD is described: at the beginning of each surgery, the user can enter patient data or select one patient from the worklist at the human machine interface (HMI) of the surgical microscope.

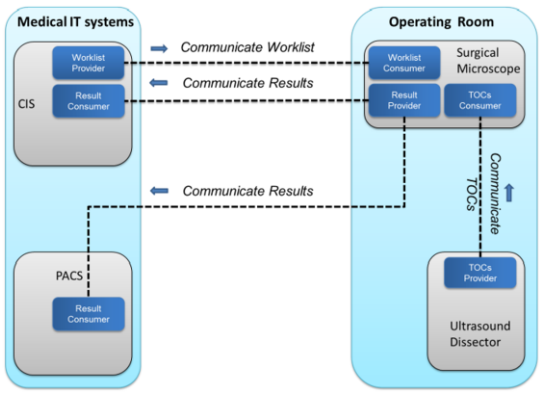


Figure 7 – Macrointegration profile for use case surgical microscope and ultrasound dissector

In the second case the worklist has to be queried from the worklist provider. This communication represents our first microintegration profile Communicate Worklist with CIS as a worklist provider and the surgical microscope as a worklist consumer. Patient data preview is supported by the HMI of the microscope. The second use case describes the triggering of USD functions, the use of USD parameters, and signaling USD error warnings over the HMI of the surgical microscope. For that the surgical microscope and the USD have to be previously configured. Using only one user interface for both devices would simplify handling both at the same time and provide better overview. For safety and security reasons, both USD parameters and functions must still be set, read, and triggered in the USD independently of any control by the HMI of the microscope. Settings on the USD always take precedence over conflicting settings on the HMI of the microscope. Using the HMI of the surgical microscope, current settings of the USD parameters can be visualized and changed on the microscope. The user can trigger USD functions over the HMI of the microscope and is warned when he exerts too much pressure with the USD instrument. Those communications are represented by the fourth microintegration profile Communicate TOCs. In this case, the surgical microscope represents the TOCs Consumer and USD the TOCs Provider. At the end of the neurosurgery, all stored patient data can be archived in medical IT systems. Picture and video recordings can be archived in a PACS and patient-relevant data are stored in the clinical information system. The microscope acts here as a result provider, and CIS and PACS act as result consumers. Thus, microintegration profiles can be combined to accomplish needed communication requirements for specific surgery and user requests.

Based on a literature search of existing IHE profiles, we identified existing IHE transactions that can map our communication needs. For each of the first three MIPs the identified IHE transactions are described below.

MIP "Communicate Worklist"

For this MIP the IHE Radiology domain provides a mechanism for communicating a DICOM Worklist in the Scheduled Workflow Integration Profile. An Acquisition Modality actor like a CT Scanner receives a DICOM Worklist after querying it from the Order Filler actor using the IHE transaction Query Modality Worklist [RAD-5]. This transaction may be used by medical devices with a DICOM interface to receive a DICOM Modality Worklist. In general DICOM interfaces are provided from medical imaging devices. One of the physicians' requirements is to provide a universal worklist via HL7 for medical devices having no DICOM interface, but having an HL7 interface. At the moment IHE provides no transaction for communicating an HL7 worklist. One mechanism to handle this missing transaction is to provide a component that listens for appropriate HL7 messages like order messages, stores the information in a local database, and answers queries from various devices.

MIP "Communicate Patient Demographics"

For this MIP we identified the IHE Integration Profile Patient Demographics Query (PDQ). Applications can query a central patient information server and retrieve patient demographics. One can query partial or complete patient name, patient ID, date of birth, bed ID, etc. The Patient Demographics Query [ITI-21] transaction is triggered by the Patient Demographics Consumer actor and addressed to the Patient Demographics Supplier actor. If visit information is also required, one can use the Patient Demographics and Visit Query [ITI-22] transaction. The Patient Administration Management (PAM) profile is an alternative to PDQ. While the transaction Patient Identity Feed [ITI-30] provides the same category of data in push mode (notification/acknowledgement), PDQ operates in pull mode (query/response).

MIP "Communicate Results"

For communicating multimedia results, the IHE Radiology transactions Query Images [RAD-14] and Retrieve Images [RAD-16] can be used. During surgery the physician can display images with those DICOM-related transactions from an Image Archive actor. Intraoperative images can then be stored at the Image Archive/Manager actor with the Modality Image Stored [RAD-8] and Storage Commitment [RAD-10] transactions.

For communicating alphanumeric results between, for example, physiological monitors, ventilators, infusion, and medical IT systems, the IHE Patient Care Device (PCD) domain provides the Device Enterprise Communication (DEC) Integration Profile. Device settings, device alarm state, and technical data from a medical device such as its battery charge status data can be transferred with the DEC profile. The Device Observation Reporter (DOR) actor sends device data to the Data Observation Consumer (DOC) actor by the HL7 based transaction Communicate Device Data [PCD-01] in compliance with IEEE 11073 terminology mapping. This allows semantic interoperability.

MIP "Communicate Technical Operation and Control Parameters" (TOCs)

For communicating TOCs between medical devices in the operating room, no IHE integration profile could be identified.

Figure 8 shows the abstract macrointegration profile after the mapping to existing IHE Integration Profiles, whereat the actors of the individual IHE transactions were adopted.

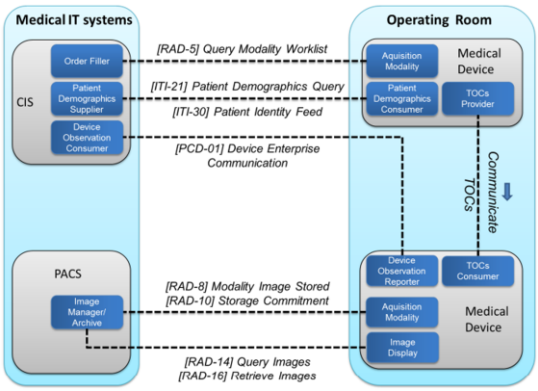


Figure 8 – Macrointegration profile with existing IHE transactions

Discussion

Since patient demographic data are essential for an upcoming surgery, they have to be automatically transferred to the operating room in an accurate way. These data can be transferred using established standards like HL7 and DICOM.

During surgery, medical devices deliver relevant data for medical IT systems. These data are mostly transferred by proprietary protocols. For communication of device data to medical IT systems, IHE provides the HL7-based integration profile Device Enterprise Communication as a first profile within the IHE Patient Care Device domain [9]. Connectathon results show that 51 vendors of medical devices and medical IT systems already implemented this integration profile, and 18 of them implementing both actors: the Device Observation Reporter and the Device Observation Consumer [10].

Since IHE does not provide integration profiles for the communication between medical devices in the operating room, this paper provides basic microintegration profiles as a first step towards new IHE integration profiles. As a next step, the identified actors and transactions have to be specified on a technical level, including communication protocols, data model, and the semantics of exchanged data.

Gathering and sharing medical device information with other devices using communication standards opens new ways in an Internet of Things (IoT) era. As IoT becomes relevant for the healthcare domain, further investigation is required in order to examine how standardizing efforts in healthcare and IoT interact.

Conclusion

This work identifies communication needs between medical IT systems and the operating room as well as between medical devices. Identified communication needs could be reduced to four abstract communication paths formulated as MIPs. Furthermore, a conceptual MAP was presented. According to the literature, the transactions of the conceptual MAP could be mapped to existing IHE transactions, except for the transaction “Communicate TOCs”.

Acknowledgments

This work is part of the OR.NET project, funded by the German Federal Ministry of Education and Research (BMBF).

References

[1] HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations. 2nd Edition. 2010. [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://www.himss.org/library/interoperability-standards/what-is-interoperability

[2] MD PnP Getting Connected for Patient Safety: Medical Device Plug-and-Play Interoperability Program [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://www.mdpnp.org/uploads/Oct12_MD_PnP_White_Paper.pdf

[3] OR.NET [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://www.ornet.org/

[4] IHE [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://www.ihe.net/

[5] DICOM [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://dicom.nema.org/

[6] HL7 [Internet]. 2014 [cited 2014 Nov 30]. Available from: http://www.hl7.org/

[7] IHE Process [Internet]. 2014 [cited 2014 Dec 04]. Available from: http://www.ihe.net/IHE_Process/

[8] R. Pahontu, A. Will, B. Bergh: Towards Communication Requirements in the Operating Room and Clinic IT, In 9th scientific eHealth conference (eHealth), 2015, accepted for publication.

[9] J. G. Rhoads, T. Cooper, K. Fuchs, P. Schluter, and R. P. Zambuto: Medical Device Interoperability and the Integrating the Healthcare Enterprise (IHE) Initiative [cited 2015 March 31]. Available from: http://s3.amazonaws.com/rdcms-aami/files/production/public/FileDownloads/HT_Interoperability/2010.ith.Interoperability_Rhoads.pdf

[10] Connectathon results [Internet]. 2015 [cited 2015 March 31]. Available from: http://connectathon-results.ihe.net/view_result.php?rows=company&columns=actor&title=integration_profile

Address for correspondence

Dipl.-Inform. Raluca Pahontu

Center for Information Technology and Medical Engineering, Heidelberg University Hospital, Germany

e-Mail: raluca.pahontu@med.uni-heidelberg.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-227

Product-based Safety Certification for Medical Devices Embedded Software

José Augusto Neto[a], Jemerson Figueiredo Damásio[a], Paul Monthaler[a], Misael Morais[a]

[a] *State University of Paraíba – UEPB, Campina Grande, Paraíba, Brazil*

Abstract

Worldwide medical device embedded software certification practices are currently focused on manufacturing best practices. In Brazil, the national regulatory agency does not hold a local certification process for software-intensive medical devices and admits international certification (e.g. FDA and CE) from local and international industry to operate in the Brazilian health care market. We present here a product-based certification process as a candidate process to support the Brazilian regulatory agency ANVISA in medical device software regulation. Center of Strategic Technology for Healthcare (NUTES) medical device embedded software certification is based on a solid safety quality model and has been tested with reasonable success against the Class I risk device Generic Infusion Pump (GIP).

Keywords:

Medical Devices; Embedded Software Certification; Safety Certification; Certification Process.

Introduction

The software embedded in medical devices introduces new features and defines competitive differences among same-class equipment. Despite this benefit, the presence of software, and its intense interaction with electrical and electronic components, implies new malfunctioning risks and potential harm to patients and healthcare professionals who interact with it. In the last decade, the US regulatory agency (FDA) has published information concerning fault cases on medical devices and the reports show 1,154,451 cases of damage, as well as the need for 5,294 recalls [1]. The same study points out that software is still largely responsible for recalls in medical devices, and is directly involved in 33.3% of recalls for Class I (high risk) equipment, 65.6% of Class II (medium risk) and 75.3% of Class III (low risk).

The FDA provides some guidance to the industry on the general principles of software validation. In a specific document [2], the FDA specifies that software must meet a set of rules on the production process, as defined in the 21 CFR §820.70 law [3]. The legislation in Brazil is less specific on medical device embedded software, but recently the Brazilian regulatory agency ANVISA exposed a new norm to public consultation [4] which, in short, proposes to comply with the FDA production-based approach.

Overall, the current certification models are mainly based on ensuring that the production processes follow international standards such as ISO 14971 [5] and ISO 80002 [6]. Promising alternatives are known, such as the argument-based approach, where manufacturers provide safety cases and certification authorities assess them individually, and the safety quality model approach, in which quality questions are provided to guide assessors to decide whether a software product is safe or not [7]. This work extends the NUTES-IESE/Fraunhofer Software Safety Quality Model [7] with a safety certification process based on it.

The Product Certifier Body (OCP in Portuguese) is the group appointed by ANVISA that have expertise to audit medical device production processes. However, to the best of our knowledge, there is no OCP accredited to perform specific certification for embedded software in medical devices, even if the focus is on the production process.

NUTES Quality Model and the Certification Process Model

The Center for Strategic Technologies in Healthcare (NUTES) was created in 2011, as result of a partnership between the State University of Paraíba (UEPB – www.uepb.edu.br) and the Brazilian Health Ministry, aiming to develop products and technologies to support the Brazilian medical device industry, as well as the regulatory agencies that supervise it.

In 2013 and 2014, the NUTES technical team, in cooperation with the German institute IESE/Fraunhofer-Kaiserlautern, developed a safety quality model for medical device embedded software totally centered on the final product, instead of the manufacturing best practices. The basis of the NUTES Quality Model is a strong theoretical framework of safety engineering, combined with the Fraunhofer background, which has been accumulated through many years of providing services to industries in which embedded software is critical, such as automotive and the airline industry [8].

The NUTES Quality Model proposes asking a set of quality questions against the software documentation; those questions precisely define what is considered safe software in a medical device [7]. In spite of that, the model does not include a process for how these questions should be asked, nor does it provide a step by step tutorial to check whether a software is safe or not.

This paper presents a structured process used at NUTES to certify embedded software in medical devices. The process is original to Brazil and is not production-based, but rather product-based. It uses the NUTES Quality Model and, as a consequence, it is reusable and adaptable to any software-intensive medical device. This is possible due to the quality model structure, based on general safety cases, which in turn refers to requirements, architecture, testing, and code aspects (common to every software product).

The certification process presented in this work can be, for didactic purposes, split into three phases: 1) the initial customer interaction; 2) the documents acquisition to support the certification realization; and 3) the software analysis. Particularly for phases 2 and 3, we list the input and output artifacts. We also present the criterion of division in 5 subareas of technical knowledge (safety, code, tests, architecture and usability), and the aggregation rules between areas, which produces a quantitative and qualitative outcome of the certification.

The provided process is compatible with the technical requirements demanded by ANVISA, and is continuously evolving to stay aligned with technical requirements

Pilot Evaluation

Manufacturers resist collaborating in experimental assessments since NUTES certification exposes classified product features. Moreover, ANVISA still does not require a local software certification. As an alternative, we performed a pilot evaluation of the process applying it to certify the Generic Infusion Pump (GIP) [9]. The GIP is a project developed by the FDA Software Engineering Laboratory in cooperation with the University of Pennsylvania.

The GIP failed the NUTES certification process, mostly because we could not satisfy part of the quality model, either by lack of documents or due to real product safety issues. At the end of this document, we present a proper discussion on the results, implications and limitations of this approach.

Methods

Theoretical Background

The process described here results from a bibliographic review of documents provided by regulatory agencies that describe the main product certification steps [10, 11]. Moreover, the NUTES certification team is working on the ISO 17025 [12] certification pipeline; and in planning the next step, the ISO 9001 [13].

The process described next is documented within a full quality management system, compliant with international standards. We built the process assisted by an ISO consulting company. The interaction between NUTES and this company was formal, and occurred during eight months in 2014, from March to November. Furthermore, there are several direct and indirect contributions of members of the FDA and ANVISA to the current version of the certification process. The latter is much closer and especially important since many ANVISA members also act as professors at NUTES Postgraduate Program in Science and Technology in Health. The result of this effort is a structured model, presented in Business Process Modeling Notation (BPMN), and a list of documents that support its execution.

We executed the GIP certification process in September 2014, in a 15-day period, and it was attended by the NUTES Software and Safety Engineering Research Team. Currently, this team is composed of PhD Computer Science faculty members with relevant contributions in the software engineering field [7, 14, 15, 16]. The team has been through a two-year safety-focused training by means of a cooperation with the experienced IESE/Fraunhofer Kaiserlautern.

Evaluation Method

We primarily evaluated the process in a pilot experiment for the GIP certification. We consider the infusion pump choice adequate for the following specific reasons:

1. The FDA claims [1] that infusion pumps are a continuous source of safety problems;
2. The GIP in an open specification. This is interesting for future evaluation and replication of the efforts in next phases of this work. Some obstacles could emerge if we had opted for a proprietary device.
3. The GIP is a robust project of a regulatory agency and its documentation works as benchmark for both manufacturers and the academic community.
4. It is a popular project, cited in more than 10 academic papers.
5. Its conception is safety-focused, therefore directly addresses the requirements of this work.

We used the GIP publicly available documentation [17] as input to the certification process; it includes articles, models and code. Complementary documentation was required and produced by means of NUTES internal efforts [18]. As a result, we present an argument-based validation and discuss the impact of the input documentation on the process.

Results

Certification Process Model

The certification process entails three distinct and sequenced phases: **Phase 1 (Business agreement)** – contains the customer-NUTES interactions up to hiring the certification. In this phase, negotiation and contract signature occur; **Phase 2 (Document submission)** – comprises the actions performed with the purpose to acquire the artifacts required to execute the certification process. This phase is interactive and incremental (i.e., we held document evaluation rounds and additional documentation is requested when necessary). The last phase takes place when the interactions end and the documentation is complete; **Phase 3 (Software analysis)** – the main certification phase. We share the documentation among the teams in charge of subareas of certification. The subarea certification is performed, and the individual results are composed into a final certification result.

The brief description that follows summarizes the knowledge recorded in 19 operational processes and 15 document templates. The big picture of this work includes the link between this certification process model and a quality management system, together they configure the minimum elements required by ISO 17025 and ISO 9001, for which NUTES is to be submitted.

Figure 1 details the certification process using BPMN. Notice that Phase 1 is not present in the flow since it is all about commercial interaction between NUTES and the product manufacturer, or ANVISA (both are potential clients for embedded software certification). This interaction includes defining the risk class of the equipment and the embedded software under certification, as well as the milestones and deadlines, specially related to the current specification of the NUTES technical team.

Composing the Certification Dossier

The certification process starts in Phase 2, in which the software documentation is requested from the client. This request specifies deadlines and delivery orders for the documentation. We analyze all documents received and depending on its completeness for the type of device under evaluation, we decide whether the received documentation meets the process needs; if it does not, we send a new request. The input software documents for certification are:

1. Test Plans – a document that guided the execution of different categories of tests in the system;
2. Test Logs – a detailed record that includes time marks of the execution tests results;
3. Test Reports – a formal document that describes the tests results, where successes and failures are presented;
4. Architectural Document – a high-level view of the system. It uses text and architectural models to

explain the system in a macro vision (e.g.: which module interacts with the alarm module);

5. Functional Specification – presents the system expected behavior;
6. Safety Requirements Document – this document lists the body of functionalities directly related to guarantee the safety of the system;
7. Software Design – it describes the system in a still abstract perspective, but in a much lower level than the architecture (e.g.: how small code unities the interaction that occurs inside the alarm module);
8. Software Source Code and Support Libraries – the system itself, coded in a known programming language;
9. Usability Engineering Document – it describes the decision behind the way the interaction between system and user is considered;
10. User Manual – a document addressed to the final user. It explains the correct way one should act in order to interact with the system;
11. Alarms Specification – it describes the events that might happen, and how they fire observable notifications to the user;
12. Risk Management File – an important document, suggested by ISO 80002 [6], which exposes every safety aspect of a device and how they are considered (e.g.: known risks, fail occurrence chance for different categories, etc.).

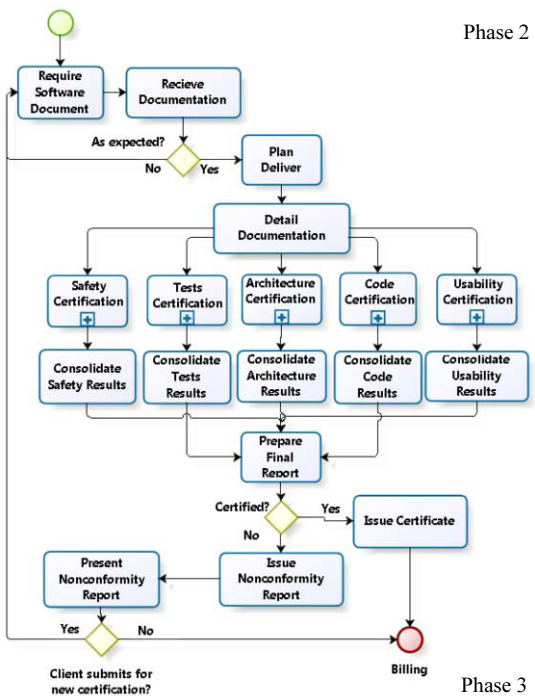


Figure 1 – The certification process for medical devices embedded software

Notice that these documents are a subset of those required by the FDA. This indicates that their production is typical, and does not impose drastic changes to the way that the software industry currently works. Moreover, in the case of certifying a real software instead of an open specification, the manufacturer is provided with "The NUTES Guide for Certification of Embedded Software for Clients". The guide goes into detail on each required document and also presents sample documents with the completeness level expected for the certification process.

Phase 2 takes at least one day and at most a week; this variation depends on the customer documentation availability. Once documentation is accepted, the third phase starts.

In Phase 3 we establish a delivery plan for each team (i.e., safety, test, architecture, code and usability) involved in the certification. This plan takes into account the certifications in progress and the teams' availability. Next, the full documentation body is detailed and shared (with possible copies) for the subareas. Finally, they generate specific reports.

Figure 2 presents the process executed in each subarea. The first task is about receiving documents and checking if the documentation is complete. At this point, any team can make requests for additional documentation to the certification manager who controls the documents originally received in Phase 2.

The subarea planning certification represents the first interaction with the NUTES Quality Model. The first step is about defining the set of questions that apply to the device under certification. At this point, teams should select questions and classify each of them as mandatory, desirable or optional according to device's risk class. In the second step, we define deadlines for the execution and report delivery. All steps, including the choice of questions and their classification, demand prior technical background and should be endorsed by a previous certification executed at NUTES.

The execution task refers to the assessment of available documentation seeking answers to the selected questions. For certification purposes, the NUTES Quality Model's questions must receive positive answers. Negative responses indicate a possible non-compliance of the software.

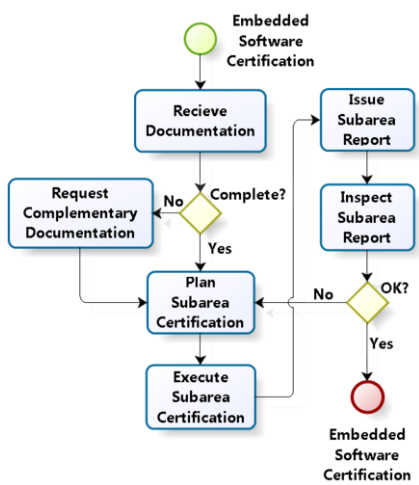


Figure 2 – The subarea certification process for medical devices embedded software.

In the end, each area produces two reports: a qualitative description of the product, and a quantitative result of the subarea certification. The first shows the assessed documentation, and presents a textual argument about the strengths and weaknesses of the product, highlighting nonconformities that led to negative answers (if any). Finally,

if the software fails the certification, the report suggests modifications to the software, such as new features to enable a better score on future certifications. The quantitative report describes the number of positive answers for each class of questions in comparison to the maximum score per class. Figure 3 shows an example of part of a quantitative subarea report. We express the number of positive and negative responses for each class in both textual and graphical representations.

Mandatory		Desirable		Optional	
Yes	No	Yes	No	Yes	No
15	0	3	2	1	1
100%	0%	60%	40%	50%	50%

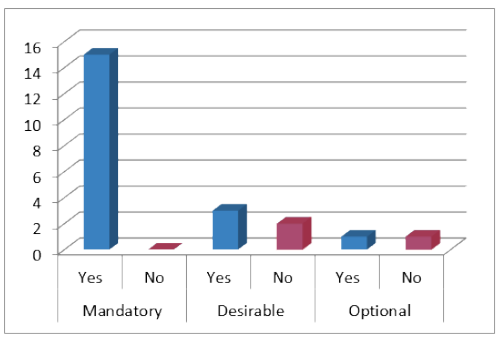


Figure 3 – Subarea quantitative results showing the total number of positive and negative answers for each class.

Finally, the team leader inspects the subarea report looking for mistakes or inconsistencies. If the team members finds any divergence, they must revisit the whole subarea process.

Returning to the main process, the certification manager analyzes each subarea report and produces the final report. The final report consists, first, of all subareas qualitative observations presented in a coherent outline. Next, it includes the subareas quantitative reports in a summarized table, with both "Yes" and "No" responses for each class of questions. To achieve the final result, the following aggregation rule is applied: questions related to mandatory safety features have a limit of zero negative answers allowed. Questions related to optional and desired safety features have an upper limit for negative answers according to the device's risk class. For the GIP, two is the limit for negative answers in desirable questions, and a varying number of negative answers is acceptable for optional safety features related questions. Each area specialist, based on international standards practices and individual expertise, defines those bounds.

Once the report is ready, one of the following situations occurs: 1) A certification is assembled and sent to the customer if it is compliant (i.e.: negative answers did not go over the limits) and therefore the embedded is considered safe; or 2) The customer is told the software failed the certification and is provided with the report that clarifies the problems that led to the negative result.

Pilot Evaluation

We applied the certification process to the GIP and the results show that the device, as currently described in the public specification, is not safe as a software intensive medical device is required to be according to NUTES Quality Model and its 300 safety-related quality questions.

However, due to confidentiality reasons, we cannot elaborate on the real results for any subarea but Safety. Such restriction does not influence the certification process comprehension, the main goal of this work. Table 1 presents the quantitative results, where only Safety results are accurate while others are slightly modified. We altered these results in a way that conclusions are the same from the original report, as follows.

Table 1 – Quantitative result of GIP certification against NUTES Quality Model questions

Subarea	Mandatory		Desirable		Optional	
	Yes	No	Yes	No	Yes	No
Usability	3	10	12	16	14	1
Safety	20	5	1	0	0	0
Architecture	5	10	17	1	2	10
Requirements	18	3	36	5	6	20
Code	15	0	4	1	1	1
Test	4	36	11	6	3	3
Total	65	64	81	29	26	35

A full "Yes" score is a must for Mandatory questions on every subarea; for the Safety subarea (real result), GIP's 20 "Yes" out of 25 questions is enough to fail the certification. The illustrative numbers used for the remaining subareas show that we address Usability, Architecture, Tests and Requirement's safety issues properly, and Tests holds the weakest safety treatment. On the other hand, Code alone had the best result, mainly because GIP's code was a good source, where we found positive answers for all questions.

Discussion

The GIP project is largely different from conventional software. Its documentation derives from multiple academic works, while conventional documentation is generally produced in sequence, and each document is clearly related to others during the software engineering process. In this paper, we have claimed that we used all the available documentation for GIP. The first problem is that the documentation found was not enough to fill out the required list. For example, we used the Hazard Analysis [19] and a document of Risk Analysis [18] to address the lack of a Risk Management File. Therefore, from the 64 negations for mandatory question, only 14 were based on real evidence. The other 50 were explained in terms of "No evidence in the documentation".

In a real product scenario, the information scarcity is also possible, but we would be able to ask the customer for extra documentation. Another aspect is the documents internally produced by NUTES team to reach the documentation required by the process, we believe that the certification scores could be higher if extra effort was placed in better composing of mock documents. Additional in-depth considerations, specifically for the Architecture subarea, can be found in another work [14].

Conclusion

The regulation for embedded software in medical devices in Brazil is currently based on international certifications, such as the FDA or CE. These certifications focus on the audit of manufacturing processes, assuming that well-executed processes will generate safe products, but this approach has problems. In this work we presented a certification process focused on the product, supported by a quality model that provides a general safety case for embedded software in medical devices.

We applied the designed process to the GIP and the results point to the high standards to which a software-intensive medical device must comply with to be considered safe, and therefore certified. We have fully executed the process, but real product certifications are necessary to better validate and improve the process proposed here. First agreements with industry partners have already taken place, and we intend to execute new rounds of the process on commercialized products soon.

The certification process is currently undergoing ISO 17025 certification, and shall be submitted to the ISO 9001 soon. A first internal audit phase for ISO 17025 has already been accomplished. This step fulfills accreditation requirements to NUTES be able to provide ANVISA with a local software certification service to Brazilian and global manufacturers.

Acknowledgements

This research is directly supported by NUTES/UEPB and several of its researchers, who have made multiple contributions in the pilot evaluation results. We also thank Dr. Kleber Nobrega and his team at Perceptum, who offered strong methodological support for us to create and refine the certification process presented here.

References

[1] Alemzadeh, H.; Iyer, R.K.; Kalbarczyk, Z.; Raman, J. Analysis of Safety-Critical Computer Failures in Medical Devices. IEEE Computer Society 2013;11:14–26.

[2] FDA, General Principles of Software Validation; Final Guidance for Industry and FDA Staff Health. San Francisco 2002; 47.

[3] FDA, 21CFR820.70. [online: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=820.70. Accessed 2014 Nov 3rd].

[4] Portaria 407, Consulta pública sobre o Aperfeiçoamento dos Requisitos de Avaliação da Conformidade para Equipamentos sob, 2014 August 26th. [online: http://www.inmetro.gov.br/legislacao/rtac/pdf/RTAC002160.pdf. accessed: 2014 Nov 16th]

[5] ISO/TC 210, Medical devices - Application of risk management to medical devices, 2010, ISO/TR 14971:2007.

[6] ISO/TC 210, Medical devices software – Part 1: Guidance on the Application of ISO 14971 to medical device software, 2009, ISO/TR 80002:2009.

[7] Adler, R.; Kemmann, S.; Filho, D. M.; Augusto, J. Safety assessment of software-intensive medical devices. Proceeding of IEEE Symposium on Software Reliability Engineering Workshops 2013;217-22.

[8] Fraunhofer Institute Research Topics - Transportation and Mobility [online: http://www.fraunhofer.de/en/research-topics/transportation-mobility.html. Access: 2014 Dec 3rd].

[9] The generic patient controlled analgesia pump model. [online: http://rtg.cis.upenn.edu/gip.php3. accessed: 2014 Nov 15th]

[10] FDA Overview of Device Regulation, 2014. [online: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/. Access: 2014 Nov 12th]

[11] Europe CE Approval Process for Medical Devices, 2014 [online:http://www.emergogroup.com/pt/resources/europe-process-chart. Access: 2014 Nov 11th]

[12] ISO/CASCO, General requirements for the competence of testing and calibration laboratories, 2010, ISO/TR 17025:2005.

[13] ISO/TC 176/SC 2, Quality Management Systems - Requirements, 2008, ISO/TR 9001:2008.

[14] Leite, F.; Antonino, P.; Barbosa, P.; Kemmann, S.; Mendonca, R. Are the Current Architectural Practices Suitable for Safety Aspects of Medical Devices? An Exploratory Investigation. IEEE HealthCom, 2014.

[15] Costa, Túlio H. ; Andrade, Rodrigo ; Maciel, Edna Queiroz; Lima, Lukas Teles ; Bublitz, Frederico Moreira . A Safety Engineering Hazard Identification for Hemodialysis Systems. In: 2015 International Conference on Consumer Electronics, 2015, Las Vegas. International Conference on Consumer Electronics, 2015. p. 86-87.

[16] Silva, L. C.; Perkusich, M.; Bublitz, F. M. ; Almeida, H.; Perkusich, A. A model-based architecture for testing medical cyber-physical systems. In: the 29th Annual ACM Symposium, 2014, Gyeongju. Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC '14. p. 25-30.

[17] The Generic Infusion Pump publications [online: http://rtg.cis.upenn.edu/gip_pub.php3. Acessed: 2014 Nov 14th]

[18] NUTES GIP Documentation [online: https://www.dropbox.com/sh/pbz4qncfokvc938/AABPRIOckiluQfzcYHH1Doepa?dl=0. Access: 2014 Dec 7th]

[19] Yi Z.; Paul J.; Raoul J.A Hazard Analysis for a Generic Insulin Infusion Pump. Journal of Diabetes Science and Technology, March 2010, 4 (2):263-83.

Address for correspondence

José Augusto de Oliveira Neto, E-mail: zedeguga@gmail.com, Baraúnas st. 351 - Bairro Universitário - Campina Grande/PB - Brazil, Zip 58429-500

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-232

Patient Safety at Transitions of Care: Use of a Compulsory Electronic Reconciliation Tool in an Academic Hospital

Daniel A. Rizzato Lede, Sonia E. Benítez, John C. Mayan III, María I. Smith, Analía J. Baum, Daniel R. Luna, Fernán González Bernaldo de Quirós

Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Medication errors are responsible for most inpatient adverse events. Medication reconciliation emerged as an effective strategy to decrease these problems, enhancing patient safety. Electronic health records with reconciliation tools could improve the process, but many aspects should be considered in order to reach expected outcomes. In this paper we analyzed how a compulsory, electronic reconciliation application was used at Hospital Italiano in Buenos Aires, through admission and discharge processes. We evaluated all medications that were reconciled during patient admission and discharge since its implementation, from February to November 2014. During that period, there were 78,714 reconciled medications regarding 37,741 admissions (2.08 reconciled medications per hospitalization), of 27,375 patients (2.88 medications per patient). At admission, 63% of medications were confirmed and the remaining were paused or deleted. At discharge, 41% of all medications were reconfirmed. In the creation of the best possible medication history, the use of an electronic reconciliation tool would clean overloaded lists, but at the same time medications could be erroneously deleted.

Keywords:
Electronic Health Record; Inpatient Data; Cross sectional study; Medication safety

Introduction

Medication errors (ME) are the main cause of inpatient iatrogenesis, followed by postoperative complications and hospital acquired infections [1,2] Up to 20% of hospital adverse drug events (ADE) take place at admission or discharge and are related to poor inter-professional or patient-professional communication [3]. More than 50% of the inpatient population has one or more unwarranted discrepancies between their prescription list and the actual, taken medications [4]. These shortfalls can lead to omission of crucial chronic medication, therapeutic duplication and names or doses mistakes. Almost 60% of these errors could potentially cause serious health risks for the patient [5,6]. Medication reconciliation (MedRec) is "a process of identifying the most accurate list of all medications a patient is taking,… and using this list to provide correct medications for patients anywhere within the healthcare system,"[7] thus, preventing ME especially during transition of care. A comprehensive achievement of the best possible medication history (BPMH) is particularly important [8].

Electronic tools could help to perform MedRec, consequently improving patient safety [9,10]. The introduction of an electronic reconciliation application showed a 28% relative risk reduction of potential ADEs [11]. However, effective implementation of MedRec has proven to be challenging, and clinical endpoints are currently subject to further studies [12,13]. There have been studies showing less than 20% adoption of optional electronic MedRec applications [14]. However, there is little information about the use of compulsory tools [15]. In 2014, Hospital Italiano of Buenos Aires (HIBA) implemented a mandatory electronic MedRec tool at patient admission and discharge. The objective of this paper was to analyze how an EHR-based mandatory reconciliation tool at admission and discharge was used, in the creation of the best possible medication history.

Methods

Study Design

A cross sectional, descriptive study, with secondary analysis of a de-identified clinical database was performed.

Population

We included all of the 1) hospitalized patients at HIBA from February to November 2014, 2) medications that were reconciled during admission and discharge processes, and 3) physicians involved in the processes.

Setting

HIBA is a non-profit health care academic center founded in 1853, with over 2,700 physicians, 2,700 other health team members (including 1,200 nurses) and 1,800 administrative and support employees. HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires and its suburban area. It has an insurance plan (Plan de Salud - PS) that covers more than 150,000 people and also coordinates insurance for another 1,500,000 people who are covered by affiliated insurers. Between 2013 and 2014, over 45,000 inpatients were admitted to the hospital, with 45,000 surgical procedures (50% ambulatory) and 3 million outpatient visits.

HIBA is a teaching hospital, with over 30 medical residency-training programs and 34 fellowship programs. There are currently 400 residents and fellows in training.

HIBA runs an in-house developed Health Information System since 1998, including clinical and administrative data. Its Electronic Health Record (EHR), called Italica, is a unified, modular, problem-oriented, and patient-centered system that works in different environments (outpatient, hospital, emergency, and home care). Italica allows computer physician order entry (CPOE) for medications and medical tests, and storage and retrieval of test results, including images through a picture archiving and communication system (PACS) [16].

Analysis Plan

For this study, we analyzed: 1) the number of inpatients, gender, age, health insurance, time since first hospital contact, primary care physician (PCP) assignation; number of hospitalizations, number of admissions per patient, length of stay, admission and discharge diagnoses; 2) the number of confirmed, paused and deleted medications at admission and discharge; 3) the age, gender and specialty of physicians, seniority, authoring of admission and discharge documents.

Admission and Discharge

Before 2014, the admission and discharge documents were filled in plain text, including medical history and medications. To start the project we reviewed the literature, finding examples of electronic applications [17-21], and we worked within an interdisciplinary team (medical informaticians and computer-user interaction specialists) to reach an enhanced prototype of admission and discharge modules. A final version, implemented in early 2014, included a multiple-step wizard to fulfill information requirements for both processes, including MedRec.

In the admission *module* (Figure 1), there is a step to verify ambulatory medications to create the best possible medication history (BPMH), by checking the patient's home medication (from EHR) with direct patient interviews.

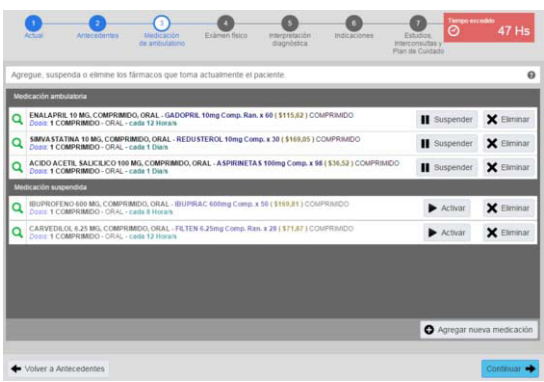


Figure 1- Admission module at EHR: the step to verify what medications the patient is taking, to create the best possible medication history.

Each prescription can be confirmed, paused, or deleted, and other drugs can be added if needed (paused medications can be reconfirmed afterwards, whereas deleted medications cannot). A subsequent step in the admissions module shows two columns to compare/reconcile the BMPH (left) and the admission medications (right), to reconcile differences. Similarly in the discharge module (Figure 2), there is a step to reconcile the prescriptions of the last day of hospitalization and the BPMH, to create the best possible medication discharge list (BPMDL). At this point, medications can be reconfirmed, paused or deleted. (See a brief tutorial at http://www.hospitalitaliano.org.ar/infomed/index.php?contenido=ver_curso.php&id_curso=17931#.VSVZpNyG_1Z)

Figure 2 - Discharge module at EHR: the step to compare medications during hospitalization (left) and ambulatory medications (right).

Results

During the 10 month period of the study, there were 37,471 hospitalizations of 27,375 patients and 205 participant physicians. With respect to patient demographics (See Table 1), the mean age of patients was 45.1 years (SD 28.57; SE 0.17). 57% of the patients were women (15,693). PS covered 41% of the inpatients (11,222); 93% of them had assigned PCPs. There were 1.36 hospitalizations per patient, most of them having just one (82%: 22,373 patients). The mean length of stay was 4.97 days (SD 8.71; SE 0.05). The most common admission and discharge diagnoses were childbirth, parturition (and C-section) and chemotherapy.

Table 1- Patient characteristics

Characteristics	**n =27,375**	
Gender		
Female	**57%**	(15,693)
Age Ranges (0-103)		
< 1 y.o.	**12%**	(3,308)
(0= newborns)	**(10%)**	(2,783)
1-20 y.o.	**11%**	(3,102)
21-40 y.o.	**22%**	(5,896)
41-60 y.o.	**17%**	(4,787)
61-80 y.o.	**26%**	(7,124)
> 80 y.o.	**12%**	(3,158)
Time since 1st hospital contact		
< 1 y	**45%**	(12,277)
1-5 y	**30%**	(8,332)
> 5 y	**25%**	(6,766)
Admissions (n=37,471)		
1	**82%**	(22,373)
2	**11%**	(3,173)
>2	**7%**	(1,829)
LOS ranges (0-194)		
< 1 d	**16%**	(6,118)
1-5 d	**58%**	(21,646)
> 5 d	**25%**	(9,199)
no data	**1%**	(508)

LOS: length of stay; y.o.:years old; y: years; d: days

Table 2 – Characteristics of physicians in charge of admission and discharge

General			Women		Men	
Characteristics		**205**	**38%**	**(78)**	**62%**	**(127)**
Specialties						
Internal Medicine	**16%**	**(32)**	53%	(17)	47%	(15)
Orthopedics	**13%**	**(26)**	12%	(3)	88%	(23)
Intensive Care	**10%**	**(20)**	50%	(10)	50%	(10)
General Surgery	**8%**	**(17)**	12%	(2)	88%	(15)
Obstetrics	**4%**	**(9)**	33%	(3)	67%	(6)
Others	**49%**	**(101)**	43%	(43)	57%	(58)
Seniority (y)						
< 1	**53%**	**(108)**	44%	(48)	56%	(60)
1-5	**30%**	**(61)**	43%	(26)	57%	(35)
> 5	**18%**	**(36)**	11%	(4)	89%	(32)
Age (y.o.)						
<35	**66%**	**(135)**	47%	(63)	53%	(72)
35-45	**19%**	**(38)**	26%	(10)	74%	(28)
>45	**16%**	**(32)**	16%	(5)	84%	(27)

y.o.:years old; y: years

With respect to physician demographics, 205 physicians were in charge of both admission and discharge documents; 38% of them were women. The majority of physicians were from Internal Medicine, Orthopedics, Intensive Care, General Surgery and Obstetrics. Two thirds of the physicians were younger than 35 years old and 83% had been in the hospital less than 5 years (See Table 2).

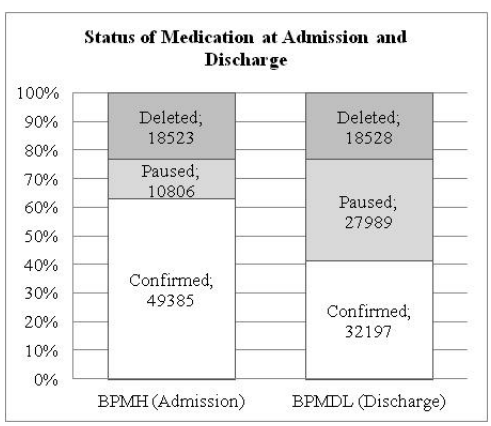


Figure 3 - Changes in medication status at transitions of care (admission and discharge)using the MedRec tool.

In the given period, there were 2.08 reconciled *medications* per hospitalization and 2.88 medications per patient. Concerning medication status, Figure 3 shows how many were confirmed, paused, or deleted at admission and discharge. The medications that were added to the BPMH using the MedRec tool were called "Newly Created" and represented just 7% of the total number of reconciled drugs (6,082 from 78,714). 58% of them were confirmed at admission (3,512), and the remaining 42% were paused or deleted (2,570). At discharge, just 44% of the newly created drugs were reconfirmed.

Discussion

The objective of this paper was to evaluate the use of an EHR-based reconciliation tool at admission and discharge. HIBA has an in-house developed Health Information System that works in both, inpatient and outpatient settings. Even though every HIBA inpatient treatment was documented using the CPOE, only PS affiliates would benefit from the electronic prescription at discharge. Electronic prescription is uncommon for other insurers and providers, who still use paper based systems. Otherwise, the MedRec electronic tool would be useful for the entire HIBA population, creating a BPMH/ BPMDL as accurately as possible.

Among the HIBA inpatient sample, almost 40% were over 60 years old. Unintended medication discrepancies are more frequent in elderly populations [22,23], related to their high number of comorbidities and medication usage, and several health profesionals' involvement [24]. The terminology data are still under further investigation, but preliminary analysis showed a significant rate of Cancer and Chemotherapy as some of the hospital's top diagnoses, in addition to maternal and neonatal health. Oncology patients are also at higher risk for medication omissions [25], and they would benefit from a MedRec program. Patients' regular length of stay was short (74% less than 5 days), according to the acute care profile of our hospital. Satisfactory MedRec could have an impact by reducing unexpected adverse drug events that could delay opportune discharge and also prevent early readmissions [26].

Our electronic MedRec tool was designed as a mandatory step in the admission and discharge process documentation. The data analysis indicated that most of these documents are completed by young and new hospital physicians, typical attributes of physicians-in-training involved in HIBA medical residency programs. Their academic training and professional responsabilities should not exclude proper supervision from experienced staff. It has been published that positive recognition programs enhance resident physician compliance with MedRec, when properly instructed and improvements are recognized publicly [27]. Although pharmacist MedRec interventions seem to be superior than other specialists [28,29], efforts should be made to engage every hospital clinician in the process, as multi-professional teams have shown better outcomes [30].

Considering the changes in medication status at transitions of care, we evidenced that it is more frequent to stop a previous prescription at discharge (pausing or deleting it) than validating it. Shortening the list could be a reasonable decision, taking into account that the most frequent unintended discrepancies would be by excess of unnecesary drugs, as shown in a previous study about medication list accuracy at HIBA [31]. Our research has limitations, some of them being related to the type of study (cross-sectional study); the database analysis (getting just part of the information from the process); and information from a single academic healthcare center, therefore the results cannot be generalized. On the other hand, thanks to this primary investigation, we are conducting new research projects to analyze prescription accuracy, and consequent impact in clinical outcomes like reported adverse drug events.

CPOE or electronic MedRec does not guarantee successful discrepancy resolution [32,33]. Many physicians may not feel prepared to perform such tasks with a compulsory tool, thus introducing new errors. We have been getting some helpdesk

incidents from PCPs asking for patients' missing chronic medication after discharge. Most of the incorrectly deleted medications were done by non-PCP physicians. We plan to conduct a mixed method study (qual-quan) to evaluate this unintended consequence. We consider it important not only to know the rate of such deletions but the ultimate reason of the deletion, to evaluate its extent and impact.

MedRec should be standardised and implemented in daily practice as a routine part of healthcare provision. EHR emerged as a useful booster for many health processes, including MedRec, but it should follow strong institutional policies and programs with regular feedback to check its outcomes.

Conclusion

This study analyzed how an EHR-based mandatory reconciliation tool at admission and discharge was used. The data showed a discontinuation dominance when editing the discharge list. Regarding the creation of the best possible medication history, our experience showed that the use of the MedRec tool would clean overloaded lists, but at the same time medications could be erroneously deleted .

Acknowledgments

This study was done under the Residency Program in Health Informatics at HIBA. We want to give special thanks to Dina Miguel, Gastón López and everyone at the Health Informatics Department who contributed to this paper. A special thanks to Patricia Yao who assisted us in proofreading this text.

References

[1] Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med [Internet]. 1991 Mar 7 [cited 2014 Jul 30];324(6):377–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1824793

[2] Aranaz JM, Aibar C, Vitaller J, Ruiz P. Estudio Nacional sobre los Efectos Adversos ligados a la Hospitalización. ENEAS; 2005.

[3] Bootman JL, Wolcott J, Aspden P, Errors CIPM, Cronenwett LR, Services BHC, et al. Preventing Medication Errors:: Quality Chasm Series [Internet]. National Academies Press; 2006. Available from: http://books.google.com.ar/books?id=87QDTqt0I-cC

[4] Tam VC, Knowles SR, Cornish PL, Fine N, Marchesano R, Etchells EE. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. CMAJ [Internet]. 2005 Aug 30 [cited 2014 Jul 30];173(5):510–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1188190&tool=pmcentrez&rendertype=abstract

[5] Gleason KM, McDaniel MR, Feinglass J, Baker DW, Lindquist L, Liss D, et al. Results of the Medications at Transitions and Clinical Handoffs (MATCH) study: an analysis of medication reconciliation errors and risk factors at hospital admission. J Gen Intern Med [Internet]. 2010 May [cited 2014 Jul 17];25(5):441–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2855002&tool=pmcentrez&rendertype=abstract

[6] Cornish PL, Knowles SR, Marchesano R, Tam V, Shadowitz S, Juurlink DN, et al. Unintended medication discrepancies at the time of hospital admission. Arch Intern Med [Internet]. 2005 Mar 28 [cited 2014 Jul 30];165(4):424–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15738372

[7] Resar R. Medication Reconciliation Review [Internet]. Boston, Massachusetts, USA; 2004 p. 1–5. Available from: http://www.ihi.org/resources/Pages/Tools/MedicationReconciliationReview.aspx

[8] Franzen K, Lenssen R, Jaehde U, Eisert A. [Medication Reconciliation-theory and practice]. Ther Umsch [Internet]. 2014 Jun [cited 2014 Dec 1];71(6):335–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24867348

[9] Agrawal A, Wu WY. Reducing medication errors and improving systems reliability using an electronic medication reconciliation system. Jt Comm J Qual Patient Saf [Internet]. 2009 Mar [cited 2014 Jul 30];35(2):106–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19241731

[10] Wong JD, Bajcar JM, Wong GG, Alibhai SMH, Huh J-H, Cesta A, et al. Medication reconciliation at hospital discharge: evaluating discrepancies. Ann Pharmacother [Internet]. 2008 Oct [cited 2014 Dec 6];42(10):1373–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18780806

[11] Schnipper JL, Hamann C, Ndumele CD, Liang CL, Carty MG, Karson AS, et al. Effect of an electronic medication reconciliation application and process redesign on potential adverse drug events: a cluster-randomized trial. Arch Intern Med [Internet]. 2009 Apr 27 [cited 2014 Jul 30];169(8):771–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19398689

[12] Salanitro AH, Kripalani S, Resnic J, Mueller SK, Wetterneck TB, Haynes KT, et al. Rationale and design of the Multicenter Medication Reconciliation Quality Improvement Study (MARQUIS). BMC Health Serv Res [Internet]. 2013 Jan [cited 2014 Dec 1];13:230. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3698100&tool=pmcentrez&rendertype=abstract

[13] Tamblyn R, Huang AR, Meguerditchian AN, Winslade NE, Rochefort C, Forster A, et al. Using novel Canadian resources to improve medication reconciliation at discharge: study protocol for a randomized controlled trial. Trials [Internet]. 2012 Jan [cited 2014 Jul 30];13:150. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3502593&tool=pmcentrez&rendertype=abstract

[14] Schnipper JL, Liang CL, Hamann C, Karson AS, Palchuk MB, McCarthy PC, et al. Development of a tool within the electronic medical record to facilitate medication reconciliation after hospital discharge. J Am Med Inform Assoc [Internet]. 2011 May 1 [cited 2014 Dec 10];18(3):309–13. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3078655&tool=pmcentrez&rendertype=abstract

[15] Bails D, Clayton K, Roy K, Cantor MN. Implementing online medication reconciliation at a large academic medical center. Jt Comm J Qual Patient Saf [Internet]. 2008 Sep [cited 2014 Dec 22];34(9):499–508. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18792654

[16] Luna D, Plazzotta F, Otero C, González Bernaldo de Quirós F, Baum A, Benítez S. Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires [Internet]. CEPAL; 2012 [cited 2014 Jul 31]. Available from: http://repositorio.cepal.org:80/handle/11362/3959

[17] Cadwallader J, Spry K, Morea J, Russ AL, Duke J, Weiner M. Design of a medication reconciliation application: facilitating clinician-focused decision making with data from multiple sources. Appl Clin Inform [Internet]. 2013 Jan [cited 2014 Jul 30];4(1):110–25. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3644819&tool=pmcentrez&rendertype=abstract

[18] Pronovost P, Weast B, Schwarz M, Wyskiel RM, Prow D, Milanovich SN, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. J Crit Care [Internet]. 2003 Dec [cited 2014 Jul 30];18(4):201–5. Available from: http://www.sciencedirect.com/science/article/pii/S0883944103001084

[19] Giménez Manzorro Á, Zoni AC, Rodríguez Rieiro C, Durán-García E, Trovato López AN, Pérez Sanz C, et al. Developing a programme for medication reconciliation at the time of admission into hospital. Int J Clin Pharm [Internet]. 2011 Aug [cited 2014 Dec 15];33(4):603–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21706311

[20] Agrawal A. Medication errors: prevention using information technology systems. Br J Clin Pharmacol [Internet]. 2009 Jun [cited 2014 Jul 30];67(6):681–6. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723209&tool=pmcentrez&rendertype=abstract

[21] Porcelli PJ, Waitman LR, Brown SH. A review of medication reconciliation issues and experiences with clinical staff and information systems. Appl Clin Inform [Internet]. 2010 Jan [cited 2014 Dec 22];1(4):442–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3633312&tool=pmcentrez&rendertype=abstract

[22] Cornu P, Steurbaut S, Leysen T, De Baere E, Ligneel C, Mets T, et al. Effect of medication reconciliation at hospital admission on medication discrepancies during hospitalization and at discharge for geriatric patients. Ann Pharmacother [Internet]. 2012 Apr [cited 2014 Dec 1];46(4):484–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22414793

[23] Van den Bemt PMLA, van der Schrieck-de Loos EM, van der Linden C, Theeuwes AMLJ, Pol AG. Effect of medication reconciliation on unintentional medication discrepancies in acute hospital admissions of elderly adults: a multicenter study. J Am Geriatr Soc [Internet]. 2013 Aug [cited 2014 Dec 10];61(8):1262–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23869999

[24] Vejar M V, Makic MBF, Kotthoff-Burrell E. Medication management for elderly patients in an academic primary care setting: A quality improvement project. J Am Assoc Nurse Pract [Internet]. 2014 May 16 [cited 2014 Dec 19]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24838763

[25] Edwards SJ, Abbott R, Edwards J, LeBlanc M, Dranitsaris G, Donnan J, et al. Outcomes assessment of a pharmacist-directed seamless care program in an ambulatory oncology clinic. J Pharm Pract [Internet]. 2014 Feb [cited 2014 Dec 19];27(1):46–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24065784

[26] Eisenhower C. Impact of pharmacist-conducted medication reconciliation at discharge on reAdmissions of elderly patients with COPD. Ann Pharmacother [Internet]. 2014 Feb [cited 2014 Dec 19];48(2):203–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24259651

[27] Neufeld NJ, González Fernández M, Christo PJ, Williams KA. Positive recognition program increases compliance with medication reconciliation by resident physicians in an outpatient clinic. Am J Med Qual [Internet]. [cited 2014 Dec 19];28(1):40–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22605782

[28] Kramer JS, Stewart MR, Fogg SM, Schminke BC, Zackula RE, Nester TM, et al. A quantitative evaluation of medication histories and reconciliation by discipline. Hosp Pharm [Internet]. 2014 Oct [cited 2014 Dec 19];49(9):826–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25477614

[29] Kripalani S, Schnipper JL. Effect of a pharmacist intervention. Ann Intern Med [Internet]. 2013 Jan 15 [cited 2014 Dec 1];158(2):137–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23318319

[30] Bondesson A, Eriksson T, Kragh A, Holmdahl L, Midlöv P, Höglund P. In-hospital medication reviews reduce unidentified drug-related problems. Eur J Clin Pharmacol [Internet]. 2013 Mar [cited 2014 Dec 19];69(3):647–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22955893

[31] Arias A, Pollán J, Gomez Saldaño AM, Luna D, González Bernaldo de Quirós F. [Concordancia entre una lista de fármacos de un registro médico electrónico y el auto-reporte de los pacientes]. In: Anais do XIII Congresso Brasileiro de Informática em Saúde [Internet]. Curitiba, Brasil; 2012 [cited 2014 Dec 19]. p. 1–6. Available from: http://www.sbis.org.br/cbis2012/arquivos/794.pdf

[32] Lee JY, Leblanc K, Fernandes OA, Huh J-H, Wong GG, Hamandi B, et al. Medication reconciliation during internal hospital transfer and impact of computerized prescriber order entry. Ann Pharmacother [Internet]. 2010 Dec [cited 2014 Dec 4];44(12):1887–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21098753

[33] Showalter JW, Rafferty CM, Swallow NA, Dasilva KO, Chuang CH. Effect of standardized electronic discharge instructions on post-Discharge hospital utilization. J Gen Intern Med [Internet]. 2011 Jul [cited 2014 Dec 15];26(7):718–23. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3138594&tool=pmcentrez&rendertype=abstract

Address for correspondence

Daniel Alejandro Rizzato Lede, MD
Health Informatics Resident
Hospital Italiano de Buenos Aires
Juan D. Perón 4190 (C1181ACH) Buenos Aires, Argentina
daniel.rizzato@hospitalitaliano.org.ar
danielrizzatolede@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-237

Lead User Design: Medication Management in Electronic Medical Records

Morgan Price[a,b] , Jens H. Weber[a,b], Iryna Davies[b] , Paule Bellwood[b]

[a] *Dept. of Computer Science, University of Victoria, Victoria, British Columbia, BC, Canada*
[b] *Dept. of Family Practice, University of British Columbia, Vancouver, British Columbia, BC, Canada*

Abstract

Improvements in medication management may lead to a reduction of preventable errors. Usability and user experience issues are common and related to achieving benefits of Electronic Medical Records (EMRs). This paper reports on a novel study that combines the lead user method with a safety engineering review to discover an innovative design for the medication management module in EMRs in primary care. Eight lead users were recruited that represented prescribers and clinical pharmacists with expertise in EMR design, evidence-based medicine, medication safety and medication research. Eight separate medication management module designs were prototyped and validated, one with each lead user. A parallel safety review of medicaiton management was completed. The findings were synthesized into a single common set of goals, activities and one interactive, visual prototype. The lead user method with safety review proved to be an effective way to elicit diverse user goals and synthesize them into a common design. The resulting design ideas focus on meeting the goals of quality, efficiency, safety, reducing the cognitive load on the user, and improving communication wih the patient and the care team. Design ideas are being adapted to an existing EMR product, providing areas for further work.

Keywords:

Electronic Medical Records; Electronic Prescription; User Interface Design; Safety; Lead User Method.

Introduction

There are an estimated 380,000 to 450,000 preventable Adverse Drug Events (ADEs) occurring annually in the United States [1]. Similarly, in Canada, there are an estimated 70,000 potentially preventable Adverse Events each year, 24% of which (approximately 16,800) are related to medication errors [2]. Electronic medication management systems are considered one way to help improve these errors.

Despite the potential of eHealth systems, such as Electronic Medical Records (EMRs), evidence of benefit has been inconsistent [3]. Designing better systems requires getting aspects "right" [4], including content, workflow, and user interfaces (UI). Designing effective UI in clinical information systems is challenging [5]. There are several UI elements to consider for clinical decision support, including: consistency, appropriate visualization, and presenting advice at the time of decision making in a way that matches clinical goals [6].

There is ongoing usability research in the area of clinical information systems; however, there appears to be a lack of design research to uncover innovative features to improve the safety and usability of the medication management within tools in systems such as EMRs. The Common User Interface project (CUI at http://www.cui.nhs.uk) was one key exception, although it focused on acute care and is no longer active.

Lead User Method

For this study, we looked for a user-centred method that promoted *abductive reasoning* about how a medication management system in an EMR *might be* designed. Design research can support this kind of abductive reasoning [7]. The lead user method is a user-centred design research method that promotes abductive reasoning by advanced users to develop product innovations [8], [9]. It fit the scope and scale of our study and we had previous experience with the method [10]. This method has been shown to lead to breakthrough products in domains outside [11], [12] and inside healthcare [13]. The goal of this method is to design innovative products [11] by engaging users who are ahead in their field of expertise and to discover their more extreme goals, needs, and how they consider meeting those needs [8], [14]. User selection is intentional and lead users are not meant to be representative: they are meant to be cutting edge in some way. The lead user method is a method that can fit into the software development lifecycle as an early requirements engineering tool and could be complemented by a number of additional approaches.

Study Objective

The objective of this study was to develop a novel method that worked with lead users to design a user interface for a medication management module for primary care EMRs.

Methods

We adapted the lead user method to clinical information systems to create a design of a medication management module for a primary care EMR. A novel contribution was to combine a safety review with the lead user design method.

Lead User Recruitment

Lead users considered for this study were: primary care clinicians using the EMR (e.g. family physicians, pharmacists in primary care) and individuals who had experience in designing EMRs or teaching or researching in the field of medication management/prescribing. Lead users were recruited through our existing network of Canadian collaborators. Potential participants received a recruitment letter and a consent form, and had an opportunity to ask questions before agreeing and signing the consent.

Lead User Design Sessions

Each lead user was invited to attend a series of three individual design sessions (held separately for each user). A research analyst (RA) was assigned to each lead user. The first semi-structured interview explored lead user's ideas related to why and how medications could be better managed through EMRs.

An interview guide was available for the RA to ensure that common workflows were considered (e.g., prescribing a new medication, renewing a medication, discontinuing a medication). Each lead user was encouraged to discuss workflows they considered important. In scope for the sessions was the use of EMRs in primary care for acute and chronic/ongoing medication management. Out of scope were other information systems such as hospital systems and personal health records. Sessions were recorded and transcribed for analysis. Goals, activities and visual design ideas were extracted from the interviews. Each lead user's requirements were translated into an interactive visual prototype using Axure RP Pro 7.0 (one prototype per lead user). The multidisciplinary research team reviewed interview findings as part of the analysis and the lead RA prototyped their user's requirements.

Follow up interviews with each lead user then focused on walking through their prototype to validate and refine their requirements. These sessions were recorded and the goals, activities, and prototype were revised between each session.

Safety Review

Safety is an important quality attribute of EMR systems. We integrated a safety review into our adaptation of the lead user method, combining two complementary hazard analysis methods: an adaptation of Failure Modes and Effects Analysis (FMEA) [15] and Hazop [16]. The methods are complementary since they provide opposite perspectives on hazard analysis. FMEA starts by enumerating accidental situations and successively identifies hazards, failure modes, and contributing events that may lead to these accidents. We used EMR/CPOE-related accident reports in the FDA's adverse event reporting system MAUDE [17] as well as input from lead users as the primary data source for defining an evidence-based set of prescribing-related accidents. Hazop analysis is performed in the opposite way; it starts with a given system design (or set of designs) and uses a systematic technique based on guidewords to identify potential workflow deviations that may lead to unsafe situations. In order to apply Hazop, we needed to define process model abstractions for the visual prototypes created as a result of the lead user design sessions. We used Keystroke Level Modeling (KLM) as the method for attaining this process abstraction [18]. The findings of both hazard analysis techniques were ranked with respect to severity, likelihood and detectability, and mitigations for the top-ranked hazards were incorporated into the final design of the medication management system.

Synthesis

The multidisciplinary research team (consisting of a physician, software engineer, computer science and health information science analysts) reviewed each lead user's design requirements and prototypes. In these synthesis sessions the analyst who worked with the lead user would act as their proxy, providing rationale for the specific design decisions. Using a combination of UI/Ux best practice and safety analysis feedback, preferred approaches were selected for the set of common goals. The synthesized requirements were captured in four ways: a goal model, an information model, an activity model, and an interactive visualization (in Axure RP Pro 7.0).

Ethics approval was received from the University of British Columbia's behavioural research ethics board (H13-01059).

Results

The full study took five months. Eight lead users were recruited for this study (nine were invited, one declined due to scheduling conflicts), none were lost to follow up during the study, and each contributed 3-5 hours of interviews. Collectively, the lead users had considerable relevant experience (Table 1) at regional, jurisdictional, national, and international levels. Each user was interviewed on average 3 times with additional email follow up with some for final validation. For each lead user, the analysis of the interviews included a list of goals and medication activities that should be supported in the EMR as well as a visual prototype validated by the lead user (analysis took two person weeks of effort per lead user).

Table 1 – Lead User Participant Characteristics

Average Age	50.5 years
Gender	4 Male / 4 Female
Family Physicians	6
Pharmacists	2
Average Years in Practice	22.5 years
Average Years using EMR	14.6 years
Involved in EMR Design	6
Involved in EMR / Rx research	6
Teaching or Academic Role	8

Safety Review

The safety review was approximately two person weeks of effort. FMEA-based hazard analysis identified 22 prescribing-related hazards, each of them associated with a list of potential failure modes and contributing factors. For example, the hazard "*no alert on harmful drug interaction*" was associated with "*alerting function expects data encoded in different coding system*" as one (of several) failure modes, and with "*hospital discharge medication encoded in different coding system*" as one (of many) contributing factors. Several assumptions were made to limit the scope of the analysis, e.g., hazards related to failed, lost or delayed communication of information over the network was declared out of scope for the analysis.

The Hazop-based analysis identified additional hazards that emerged from the specific UI design choices made in the visual prototypes. For example, one lead user design incorporated a patient-specific "news feed" that would alert the provider about any change in the patient's health process since her last visit with the provider (e.g., a patient had seen another provider who changed her medications). The source of the feed may be external, such as a health information exchange. The initial design did not differentiate between the state of "no news" and the state of the "news feed" (temporarily) being unavailable (an empty news display was used for both states). This design choice created a hazard associated with a provider confusing the state of "no news" with "no available news". Other hazards identified during Hazop analysis were related to the structure and timing of user processes (e.g., race conditions between the user entering and submitting her prescription and the decision support processing and raising alerts).

Synthesized Design:

Goals

Developing the synthesized design took four person weeks of effort. There were four main goals for an effective medication management system. The system should:

1. **Improve Quality of Medication Management**. This included reducing potential harm caused by medications through software and workflow design solutions.
2. **Improve Efficiency when Managing Medications**. This included supporting common workflows with elegant solutions that are quick for the users.
3. **Reduce Cognitive Load of the User.** This included ensuring that appropriate information is accessible when

making decisions throughout the encounter, and having the EMR data help when making choices.

4. **Improve Communication with the Patient and the Care Team.** This included providing features that help the user communicate about medications and changes clearly with the patient and the rest of the care team, even in EMRs that are not electronically connected.

Activities

Lead users described several activities that needed to be supported by the software. Through the research team's synthesis and review, the following activities were considered key (others were also prototyped).

Document and See Current Medications: users needed the ability to document a complete, single set of *current* medications for the patient. This included medications that were prescribed in the EMR as well as medications/supplements not prescribed or prescribed elsewhere (e.g., over the counter or prescribed by other providers). This included documenting instructions, indications, targets, expected end date, and current prescriber. The current medication list needed to show several pieces of additional information related to each medication, including: current instructions, current prescriber, prescription information, and presence of alerts related to that medication. Further, the list needed to be sorted/grouped by: alphabetical order, indications, prescriber, regular vs. PRN medications, and long-term vs. short-term medications.

Prescribe New Medications: Users wanted to ensure the system enabled them to prescribe new medications quickly and safely. When prescribing, users needed evidence-based guidance including support of medication choice and alerts for contraindications. Tools to support quick prescribing included leveraging recent medications (recent for the user) and user / clinic level medication favourites that would quickly populate all instructions and prescription information.

Represcribe Existing Medication(s): Many medications for chronic diseases are represcribed with their instructions unchanged. The system should support quick review of medications that need represcribing. For example, review: alerts, instructions, who the most recent prescriber is and when the prescription is expected to run out. The workflow should also allow for medications to be represcribed all at once or in batches (e.g. all antihypertensive medications, then all COPD medications), allowing for encounters to flow naturally.

Modify Medication Instructions: Over time, medication instructions may change for a patient. Medications may be titrated up or down. Modifying includes several options including: changing instructions (e.g., dose), stopping, discontinuing (with reason), and restarting a previously expired medication.

Receive and Review Decision Support (DS): Users simultaneously considered decision support to be important and often too intrusive. Lead users had several ideas where DS was needed and could be effective, including: selecting new medications, seeing the medication list, and reviewing or changing medications. There were several types of decision support that were desirable, including: medication selection support, drug alerts (e.g., interactions), and financial decision support (cost).

Lead users considered the issues related to dismissing alerts and how they may impact the patient in the future. An approach that was supported through the safety review was to allow alerts to be acknowledged but still be present/visible to the user (or another user) in the future.

Users also needed to seek information, such as requesting concise and trusted information about a medication (e.g., dose, indications, side effects) and suggesting medications or alternates in case of a contraindication.

Detailed Review of Current Medication(s): Several lead users were interested in a two-step process of review. A core set of information would always be available when seeing the current medication list (see above). A more detailed set of information would be available when the user wanted to review one or more medications. Details included the history of the instructions, specific prescriptions for the medication, and estimations of compliance. Reviewing a medication would provide information such as indicators/targets and would relate dosing history to effect; it would also provide details on alerts.

There were additional activities that were considered but not illustrated such as managing favorites.

Information Model

Figure 1 highlights the synthesized conceptual information model that is "medication centric". It includes core elements related to the medications, prescriptions, and related knowledge bases required to support medication activities. A medication can have multiple *instructions* over time and a medication can have zero to many *prescriptions*; this supports the need to review a medication's history effectively. Prescriptions can be collected in a *prescription order set* to allow for prescribing multiple medications for a patient within an encounter. Knowledge bases provide DS. To support efficiency, reusable medication *favourites* are available.

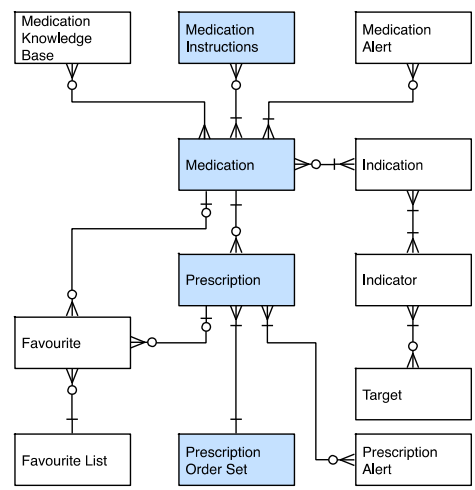


Figure 1 – The conceptual medication information model, based on the design requirements from lead users.

Synthesized Design: Visualization

The synthesized EMR UI prototype developed in this study captured the goals and activities of the lead users. The medication module UI design consists of four main panels: *current medication list*, *prescription panel, prescription order set*, and the *side panel* (Figure 2). The *EMR and patient banner* is included as misidentification is a significant safety hazard.

To enable many of the lead users' requirements, a *Medication Widget* was developed (Figure 3 and Figure 4). The widget has two states: closed and open. Each closed widget shows the medication, its current instructions, alerts (if present), and most recent prescribing information. A set of closed widgets forms the current medication list panel. The closed widget can be clicked to select the medication (Figure 5). The widget can be opened to allow the user to perform several actions on that medication: update instructions, review details (including alerts), review history, or discontinue the medication.

An *Add Medication Panel* is available when a new medication is being added. It provides the ability to add any medication (without needing to prescribe it) and provides decision support such as indications and interaction alerts. When searching for medications (by brand name or active ingredient), only active ingredient would be displayed to reduce cognitive load. Along with the formulary medications, favorite medications/prescriptions would be displayed for the user.

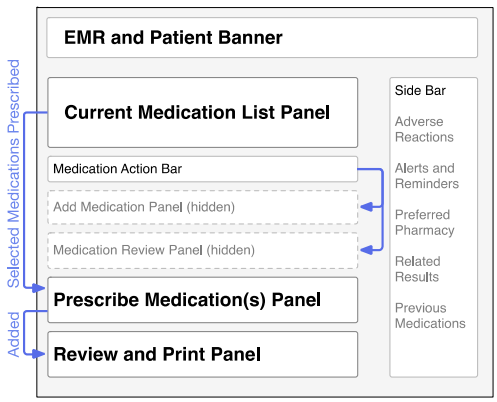


Figure 2 – Medication Management Module consists of the list of current medications, a prescribing panel, and a review and print panel.

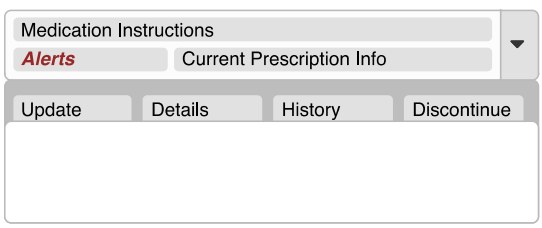


Figure 3 – The Medication Widget framework. Each widget has editing and review functionality embedded within.

Figure 4 – Example medication widet with ramipril as the current medication having one outstanding alert.

Design Comparison

The synthesized design generated through the lead user method was compared to two existing, open electronic medical records: OSCAR (www.oscarmcmaster.org) and openEMR (www.open-emr.org). We compared our design concepts to available online demo sites and user manual documentation for the equivalent core medication management functions. The new design appears to address several gaps that we saw in these products, for example: (1) a medication-centric information module (instead of a prescription module) with more coded content, (2) more robust decision support features such as suggesting alternatives, linking to indications, and having alerts visually available both when prescribing and when reviewing medications, (3) a detailed view of a medication that includes estimated adherence information and dose adjustments. However, the lead user synthesized design is currently more complex, which may prove to be challenging to new or typical users. This complexity needs to be considered and balanced when working with more typical users in busy clinics.

Discussion

This study was established to explore methods to develop innovative design solutions for medication management modules in primary care EMRs. The lead user method was adapted and applied to clinical information systems as further innovations in this area could result in significant improvements in care quality [19]. Our synthesis included the integration of a safety analysis and incorporated concepts from the literature, such as findings from the NHS' CUI project.

The lead user method was an effective way of developing the medication management module design relatively quickly. The lead user method was well received by expert clinicians. The visual validation allowed for refinement and a deeper discussion of each lead user's ideas. The synthesis by an interdisciplinary team reduced the risk of idiosyncratic design ideas from single, vocal users defining requirements in a committee or focus group. The users recommended features that were consistent with recommendations on CDSS design such as: consistency of design, appropriate visual presentation of data, use of coded data, and presenting advice at the time of decision making in a manner that matches to clinical goals [6]. Lead users have become unsolicited champions of the design, with at least half of the participants advocating for the design ideas in other EMRs or arranging opportunities to share the findings.

A challenge with this method was user selection. We found it important for lead users to have exposure to more than one EMR as well as been actively involved in the use of the EMR for research/teaching. This helped users to think outside of their own "EMR box".

Lead users thought of the EMR not just as an electronic version of the paper record, as can sometimes happen [20]. Rather, lead users considered the EMR as an interactive tool that could help with several goals: 1. Improved quality of care, 2. Improved efficiency of common workflows, 3. Reduced cognitive load, and 4. Improved communication. Decision support was an important aspect that was discussed in many areas of the medication management module and various activities. Users consistently agreed that decision support was important and they wanted it with minimal disruption. The strength of alerts was varied, depending on the workflow. Alerts would rarely stop the user from proceeding but display close to where they were active on screen (e.g., right below where they added a medication). Additional decision support such as access to knowledge bases with information on medication indications, side effects, and costs was important.

The synthesis activity allowed for integration of multiple perspectives, including the safety review and the literature. The early safety review allowed the analysis to consider lead user design ideas and potential impacts on safety and to incorporate safety elements into the design.

Limitations and Future Work

As a first phase in a design research project, further empirical evidence is needed to confirm that the design will lead to improvements in medication management. Future work is being considered along two paths: refining the methods and testing and refining the design. Three activities are proposed related to refinement and testing the design: (A) Usability testing of the prototypes with novel users. Usability testing would provide insight into the ease of use and intention to use. It was decided that the lead users would not be preferred for usability testing as (i) they were involved in the design and (ii) they were intentionally not representative of typical users. Comparative testing between multiple designs could be considered that could also show medication management errors that were

more or less likely to occur. (B) Expansion of the UI design to cover additional features. (C) Implementation of the design in an EMR. Feasibility of an implementation has been explored with the OSCAR EMR.

This study presents the results of adapting and applying the lead user method to designing aspects of clinical information systems. In this presented example, it was successful in developing a synthesized design for a medication management module for primary care EMRs. Safety was considered explicitly through the process to reduce risks to common medication errors. The resulting design included features to support quality, improve efficiency, reduce cognitive load, and improve communication. The synthesized design has been shared with OSCAR EMR and they are considering incorporating it into a future release. Future work will focus on incorporating user testing and integration into an EMR product.

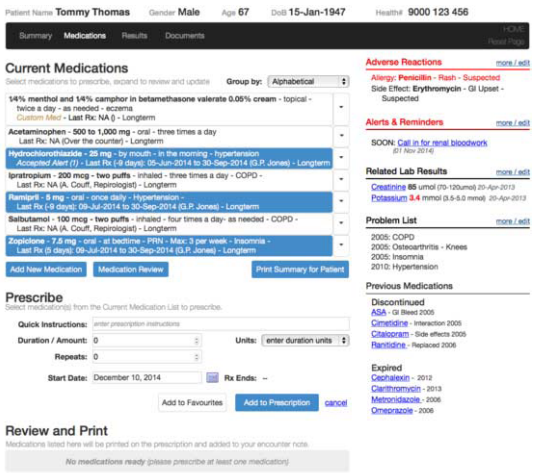


Figure 5 – The UI prototype of a medication module showing a fictitious patient with seven medications, three of which are ready to represcribed (clicked to toggle, in blue).

References

[1] P. Aspden, J. A. Wolcott, J. J. Bootman, and L. R. Cronenwett, *Preventing Medication Errors: The Quality Chasm Series*. National Academies Press, 2006.

[2] G. R. Baker, et al. "The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada," *CMAJ*, vol. 170, no. 11, pp. 1678–1686, May 2004.

[3] F. Lau, M. Price, J. Boyd, C. Partridge, H. Bell, and R. Raworth, "Impact of electronic medical record on physician practice in office settings: a systematic review," *BMC Med Inform Decis Mak*, vol. 12, no. 1, p. 10, 2012.

[4] A. B. McCoy, et al., "Clinical decision support alert appropriateness: a review and proposal for improvement.," *Ochsner J*, vol. 14, no. 2, pp. 195–202, 2014.

[5] D. F. Sittig, A. Wright, J. A. Osheroff, B. Middleton, J. M. Teich, J. S. Ash, E. Campbell, and D. W. Bates, "Grand challenges in clinical decision support," *JBI*, vol. 41, no. 2, pp. 387–392, Apr. 2008.

[6] J. Horsky, G. D. Schiff, D. Johnston, L. Mercincavage, D. Bell, and B. Middleton, "Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions," *JBI*, vol. 45, no. 6, pp. 1202–1216, Dec. 2012.

[7] T. Faste and H. Faste, "Demystifying 'design research': Design is not research, research is design," presented at the Proceedings of the IDSA, 2012.

[8] E. Von Hippel, "Lead Users: A Source of Novel Product Concepts," *Management Science*, vol. 32, no. 7, pp. 791–805, Jul. 1986.

[9] C. Lüthje and C. Herstatt, "The Lead User method: an outline of empirical findings and issues for future research - Lüthje - 2004 - R&D Management - Wiley Online Library," *R&D Management*, 2004.

[10] P. Bellwood and M. Price, "Designing electronic medication reconciliation for patients: the lead user method.," *Stud Health Technol Inform*, vol. 208, pp. 66–71, 2015.

[11] G. L. Lilien, P. D. Morrison, K. Searls, M. Sonnack, and E. V. Hippel, "Performance Assessment of the Lead User Idea-Generation Process for New Product Development," *Management Science*, vol. 48, no. 8, pp. 1042–1059, Aug. 2002.

[12] V. Bilgram, A. Brem, and K.-I. Voigt, "User-Centric Innovations In New Product Development — Systematic Identification Of Lead Users Harnessing Interactive And Collaborative Online-Tools," *Int. J. Innov. Mgt.*, vol. 12, no. 3, pp. 419–458, Sep. 2008.

[13] S. G. S. Shah and I. Robinson, "Benefits of and barriers to involving users in medical device technology development and evaluation," *J. of Inter. Tech. of Health Care*, vol. 23, no. 1, Jan. 2007.

[14] I. Eisenberg, "Lead-User Research for Breakthrough Innovation," *Research-Technology Management*, vol. 54, no. 1, pp. 50–58, 2011.

[15] P. Bonnabry, C. Despont-Gros, D. Grauser, P. Casez, M. Despond, D. Pugin, C. Rivara-Mangeat, M. Koch, M. Vial, A. Iten, and C. Lovis, "A risk analysis method to evaluate the impact of a computerized provider order entry system on patient safety." *JAMIA,* vol. 15, no. 4, pp. 453–460, Jul. 2008.

[16] T. A. Kletz, "Hazop—past and future," *Reliability Engineering & System Safety*, vol. 55, no. 3, pp. 263–266, Mar. 1997.

[17] J. H. Weber-Jahnke, "A preliminary study of apparent causes and outcomes of reported failures with patient management software.," presented at the Proceedings of the 3rd Workshop on Software Engineering in Health Care, 2011, 2011 ed., pp. 5–8.

[18] B. E. John and D. E. Kieras, "The GOMS Family of User Interface Analysis Techniques: Comparison and Contrast," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 4, pp. 320–351, 1996.

[19] E. Ammenwerth, P. Schnell-Inderst, C. Machan, and U. Siebert, "The effect of electronic prescribing on medication errors and adverse drug events: a systematic review," *JAMIA* vol. 15, no. 5, pp. 585–600, 2008.

[20] M. Price, A. Singer, and J. Kim, "Adopting electronic medical records: are they just electronic paper records?," *Can Fam Physician*, vol. 59, no. 7, pp. e322–9, Jul. 2013.

Address for correspondence

Morgan Price. Medical Science Building, University of Victoria
PO Box 1700 STN CSC, Victoria, B.C. V8W morgan@leadlab.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-242

High Override Rate for Opioid Drug-allergy Interaction Alerts: Current Trends and Recommendations for Future

Maxim Topaz[a,b], Diane L. Seger[a,c], Kenneth Lai[c], Paige G. Wickner[d], Foster Goss[e], Neil Dhopeshwarkar[a], Frank Chang[a], David W. Bates[a,b], Li Zhou[a,b,c]

[a] *Brigham and Women's Hospital, Boston, MA, USA*
[b] *Harvard Medical School, Boston, MA, USA*
[c] *Clinical Informatics, Partners eCare, Partners Healthcare System, Boston, MA, USA*
[d] *Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA, USA*
[e] *University of Colorado, Department of Emergency Medicine, Aurora, CO, USA*

Abstract

This study examined trends in drug-allergy interaction (DAI) alert overrides for opioid medications – the most commonly triggered alerts in the computerized provider order entry (CPOE). We conducted an observational analysis of the DAI opioid alerts triggered over the last decade (2004-2013, n=342,338) in two large academic hospitals in Boston (United States). We found an increasing rate of DAI alert overrides culminating in 89.7% in 2013. Allergic reactions included a high proportion (38.2%) of non-immune mediated opioid reactions (e.g. gastrointestinal upset). The DAI alert override rate was high for immune mediated (88.6%) and life threatening reactions (87.8%). Exact allergy-medication matches were overridden less frequently (about 70%) compared to non-exact matches within allergy groups (over 90%). About one-third of the alert override reasons pointed to irrelevant alerts (i.e."Patient has tolerated the medication before") and 44.9% were unknown. Those findings warrant further investigation into providers' reasons for high override rate. User interfaces should evolve to enable less interruptive and more accurate alerts to decrease alert fatigue.

Keywords:

Decision Support Systems; Clinical Drug Hypersensitivity/ Prevention & Control; Drug Therapy, Computer-Assisted; Drug-Related Side Effects and Adverse Reactions; Hospital Information Systems; Medical Records Systems, Computerized; Reminder Systems.

Introduction

During the past several decades, computerized provider order entry (CPOE) systems have become a standard and are now required for each medication order in many countries (e.g. Meaningful Use requirement in the US) [1]. These systems are vital in providing better quality care and reducing medical errors [2]. One of the important features of CPOE systems is to help providers avoid prescribing medications their patients are allergic to. This is done through drug-allergy interaction (DAI) alerts in the electronic health records. However, constant interruptive DAI alerts might overwhelm providers and cause alert fatigue – a common phenomenon where clinicians become desensitized and tend to delay or miss the response to the alarms [3,4]. There is only limited evidence on the issues related to design and implementation of DAI alerts. This study examines the rates and reasons for DAI alert overrides for one of the most common types of alerts – opioid medications.

Background

Previous evidence suggests that DAI alerts involving opioid analgesics – or narcotics – are among the most common types of alerts encountered by health practitioners [5-7]. However, true immune mediated reactions to opioids are rare and most of the opioid related reactions are non-immune mediated sensitivities or opioid induced adverse events [8,9]. True immune mediated reactions to opioids are often IgE mediated or T-cell mediated and include hives, rash, severe hypotension, bronchospasm, angioedema, and anaphylaxis [8,9]. Non-immune mediated or mimic immune reactions to opioids are usually caused by endogenous histamine release from the mast cells and cause urticaria, flushing, sweating, vasodilation, and bronchoconstriction [8,9]. Such reactions to opioids are often idiosyncratic; they may or may not recur when patients take the same opioid medication again. Other non-immune mediated side effects include nausea/vomiting, constipation, sedation, delirium, respiratory depression, and urinary retention [8,9]. In CPOE systems, a significant proportion of opioid DAI alerts are not generated by an exact match between the allergy and prescribed medication (e.g. Codeine→Codeine), but rather represent a possible cross-reactivity association between medications in similar drug families (e.g. Codeine→Oxycodone). For these reasons, opioid DAI alerts are often non-specific and might potentially increase "alert fatigue" among providers.

In the past, studies have shown that DAI alerts are frequently overridden by providers [10]. For example, two previous studies from Brigham and Women's Hospital in Boston have identified an increasing DAI alerts override trend since introduction of CPOE system over two decades ago. The overall override rates increased from about 50% in 1995 [5] to about 80% in 2002 [6]. Also, studies have shown that providers are less likely to override allergic reactions generated by an exact prescribed medication-allergy match [6, 7]. In this paper, we aimed to follow-up on this previous work and provide an update on the trends in opioid DAI alert overrides in the past 10 years. Our goals were to: 1) describe

prevalence rate of opioid DAI alert overrides 2) examine the nature of opioid-related allergic reactions (immune mediated vs. non-immune mediated; potentially life threatening vs. non-life threatening) and 3) examine factors associated with providers' tendency to override opioid DAI alerts (exact drug-allergy match vs. cross reactivity vs. drug class).

Methods

For this observational study we used data from two large academic medical centers (Brigham and Women's Hospital and Massachusetts General Hospital) from Partners Healthcare, an integrated healthcare system in the Boston area. The data generated during 10 years (2004-2013) was extracted from Partners Enterprise-wide Allergy Repository (PEAR), a longitudinal allergy database shared within the Partners provider/hospital network [11].

In PEAR, medications are encoded using a combination of local (Partners Master Drug Dictionary) and proprietary (First Databank, Inc.™) terminologies. For each patient admitted, providers are required to enter allergy information or indicate no known allergies. The recorded information combines both patient and provider reported reactions. Allergic reactions are also entered as structured or unstructured (free text) information for each allergen. Every prescribed medication is automatically checked against possible allergies in PEAR.

Drug-allergy interaction alerts are triggered when a prescribed medication matches stored allergy information in either: 1) an exact manner (definite match between main medication ingredient and allergen, i.e. Codeine→Codeine) or 2) a probable manner (probable match where allergen matches allergy group of one or more of the prescribed medication ingredients, i.e. Codeine→Oxycodone). DAI alerts present the provider with several information components, including: the allergy; the allergen; nature of the allergy-allergen match (definite/probable); and patient's reaction(s), if known. When presented with the drug-allergy interaction alert, provider needs to decide whether to cancel the medication or to override the alert. Several structured dropdown options (i.e. "Patient has tolerated drug previously") are available for the override reason together with a free text option. Providers decisions and alert information are then stored in PEAR for future reference.

Some allergy-reaction combinations are overridden less often than others (i.e. allergies resulting in anaphylaxis vs. nausea). To account for this, our research team has classified allergic reaction information into two categories based on the literature review [8,9] and consultations with allergy specialists: 1) likely immune mediated (i.e. rash or anaphylaxis) or non-immune mediated (i.e. nausea or vomiting) and 2) potentially life threatening (i.e. anaphylaxis or bronchospasm) or non-life threatening (i.e. gastrointestinal upset or nausea).

All data including allergies, allergens, medication groups, and DAI alert override rates/reasons was kept and managed using Microsoft SQL Server Management Studio [12]. For aim 1 of this study, we queried the PEAR database and summarized the frequencies of opioid DAI alerts by year. Aim 2 was achieved by summarizing the frequencies of life threatening and immune-mediated allergic reactions. For aim 3, we first queried PEAR database for medication-allergen match type (definite/probable). We then compared the frequencies of DAI alert overrides for immune-mediated and life threatening reactions by allergy-medication match type. Statistical procedures included t-tests and chi-square tests, when appropriate. The statistical analysis and group comparisons were conducted in STATA v.11 [13]. Lastly, we summarized the frequencies of the DAI alerts override reasons recorded by the providers.

This study was approved by Partners Institutional Review Board (IRB).

Results

Aim 1: Describing the prevalence and rates of opioid DAI alert overrides

Overall, the PEAR database included about one million drug-allergy interaction alerts between 2004-2013. Of these alerts, 37.3% (n=355,179) were generated for the allergens that belong to "Analgesic Narcotic Agonists" drug class (e.g. Codeine, Oxycodone, Morphine, Hydromorphone, Tramadol). This was the most common class of drugs generating drug-allergy interaction alerts in our database. While the general override rate was 83.9% for all medications, opioid DAI alerts were overridden in 88.8% cases. Figure 1 presents the increasing proportion of overrides over the years, from 85.1% in 2003 to 89.7% in 2013. We also identified a constant increase in the overall number of opioid DAI alerts from 10,500 in 2004 to 55,048 in 2013.

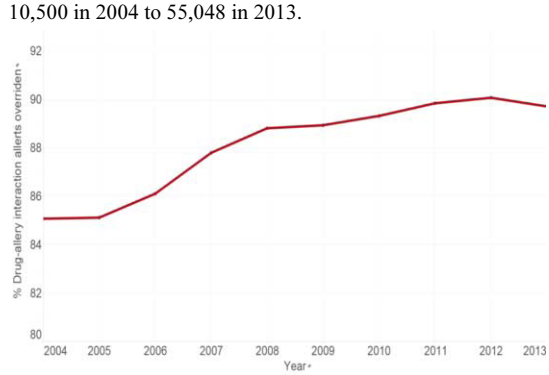


Figure 1 - Percentage of drug–allergy alert overrides for opioids over the last decade

Aim 2: Examining the nature of opioid-related allergic reactions

Table 1 presents the frequencies of the 15 most common opioid allergy reactions (covering 98% of all the reactions). Each reaction is also classified into likely immune mediated and potentially life threatening categories. 40.6% of the reactions were likely immune mediated (38.2% non-immune mediated, and the rest 21.1% unknown or not available) while only 10.4% were potentially life threatening.

Table 1 – Opioid allergy reactions

Reaction	Freq.*	%	Likely immune-mediated	Potentially life threatening
Hives or Rash	75,528	20.8	Yes	No
Gastrointestinal Upset	63,583	17.5	No	No

Itching	32,291	8.9	No	No
Mental Status Change	22,669	6.2	No	No
Vomiting	22,555	6.2	No	No
Nausea	21,820	6.0	No	No
Anaphylaxis	15,953	4.4	Yes	Yes
Shortness of Breath	5,175	1.4	Yes	Yes
Swelling (unspecified location)	4,725	1.3	Yes	No
Headaches	4,342	1.2	No	No
Angioedema	3,269	0.9	Yes	Yes
Hypotension	2,785	0.8	Yes	No
Bronchospasm or Wheezing	2,428	0.7	Yes	Yes
Seizures	1,829	0.5	No	Yes
Unknown/ unavailable	76,861	21.1	n/a	n/a
Total	**363,557**		**147,780**	**37,678**

*The total frequency of allergic reactions is larger than unique DAI alerts triggered since each unique DAI alert could have more than one associated reactions.

Aim 3: Examining factors associated with provider's tendency to override opioid drug-allergy interaction alerts

First, we examined whether there was a difference in override rates for previously reported immune mediated and life threatening reactions versus non-immune mediated and non-life threatening reactions. Although the DAI alerts override rates were statistically different, the clinical difference was only marginal for either immune mediated reactions (immune mediated overrides 88.6% vs. non-immune mediated overrides 89%, p<.001) and for the life threatening reactions (life threatening 87.8% overrides vs. non-life threatening 89% overrides, p<.001).

Further investigation revealed that majority of DAI alerts were triggered by the probable match medications (allergen matches the allergy group of one or more of the prescribed medication ingredients, 87% of DAI alerts) rather than definite match. To account for this, we conducted further comparisons based on allergy-medication match status.

We found that providers were significantly more likely to override probable matches for both immune mediated (90.7%) and non-immune mediated (91.3%) reactions whereas definite matches were overridden less frequently (72.4% immune mediated, 75.8% non-immune mediated). Similar differences (Table 2) were discovered based on life threatening reaction status which were the least overridden combination (69.9% overrides). All comparisons were statistically significant at p<.001.

In 44.9% of DAI alert overrides, we found no override reason entered by the practitioner (Table 3). Among the provided override reasons, 29% indicated previous tolerance or no allergy to the prescribed medication and 7.1% indicated that there was no reasonable alternative (provider will monitor for reaction). The remaining 17.2% of the reasons were presented as other free text with entries such as "OK with patient" or "Pain protocol". Since one-third of the cases were indicated as either previous tolerance or no existing allergy, we compared the rates of immune mediated and life threatening reactions override without those two override groups. Override trends (not shown) were similar to those presented in Table 2.

Table 2 – Drug-allergy interaction overrides by definite and probable allergen-medication matches

Reaction type	Allergy/ Medication match	Medication canceled N(%)	Alert Overridden N(%)	Overall N(%)
Immune mediated reactions to opioids				
Likely NON-immune mediated	Definite	5454 (24.2)	17103 (75.8)	22557 (6.4)
	Probable	11442 (8.7)	120290 (91.3)	131732 (37.1)
Likely immune mediated	Definite	4048 (27.6)	10634 (72.4)	14682 (4.1)
	Probable	10251 (9.3)	100196 (90.7)	110447 (31.1)
Unknown	Definite	2298 (26.1)	6514 (73.9)	8812 (2.5)
	Probable	6349 (9.5)	60600 (90.5)	66949 (18.8)
Life threatening reactions to opioids				
Likely NON-life threatening	Definite	8301 (25)	24945 (75)	33246 (9.4)
	Probable	18335 (8.8)	189961 (91.2)	208296 (58.6)
Likely life threatening	Definite	1194 (30.1)	2775 (69.9)	3969 (1.1)
	Probable	3317 (10)	29802 (90)	33119 (9.3)
Unknown	Definite	2305 (26.1)	6531 (73.9)	8836 (2.5)
	Probable	6390 (9.4)	61323 (90.6)	67713 (19.1)
Total		39842 (11.2)	315337 (88.8)	355179

Table 3 – Drug-allergy interaction alerts override reasons

Override reason	Freq.	%
Patient has tolerated drug previously	91,548	29.0
Patient reports no allergy	5,666	1.8
No reasonable alternative – will monitor for reaction	22,496	7.1
Unknown	141,437	44.9
Other (allows user to enter free text)	54,190	17.2
Overall	331,669	

Discussion

Our findings confirm a concerning trend of continuously increasing drug allergy alert overrides over the past few decades. Since the introduction of DAI alerts as a core feature of CPOE, the override rates increased from about 50% in 1995 [5], to about 80% in 2002 [6], culminating in the current rate of about 90% in 2013 in our study. We also identified an overall four-fold increase in the number of opioid DAI alerts over time. To our best knowledge, no major changes were made to the CPOE system during the study time period. This finding is consistent with the recent studies reporting a constant increase in the opioid drug prescription rates over time in inpatient and outpatient settings. For example, several recent studies have reported that opioid prescriptions increased as high as three-fold in the past two decades in the US [14-17]. It is possible that the increase in the DAI override rates is associated with an increase of opioid prescriptions. Also, opioid DAI alerts are the most common type of alerts, and providers might override more of them due to alert fatigue – a well-known phenomenon where providers become overwhelmed with the constant interruptive alerts and override more of them over time [3,4,10].

Further investigation into the nature of the opioid-related reactions has also reveled several further concerning trends. First, only about 40% of the indicated reactions were likely a result of immune-mediated processes (i.e. IgE mediated or T-cell mediated reactions such as anaphylaxis or rash). Another 40% of the reactions were likely opioid side effects, such as gastrointestinal upset or nausea and the remaining reactions were unknown or unavailable. In our study, both types of reactions were overridden in approximately 89% of the cases. Presenting accurate information has the potential to increase users buy-in to the system [18,19]. However, in our case, an inaccurate mix of allergy and side effects-related reaction information may have increased the rate of DAI alert overrides. In addition, providers might need to pay more attention to capturing reliable allergic reaction information to avoid further inconsistencies. Potentially, clinical decision support tools might help in identifying and presenting the non-immune mediated reactions differently, either at the point of allergic reaction entry by the provider or when the DAI alert is presented [19].

In an attempt to better understand providers' decision making when overriding DAI alerts, we also grouped reactions into potentially life threatening reactions versus non-life threatening reactions. Although only one-tenth of the reactions were potentially life threatening, we were surprised by the high rate of overrides. Here again, the high DAI override rates for life threatening and non-life threatening reactions were almost identical. This again underlines a critical need in decreasing the alert fatigue [4,7] and further exploration of users' experiences.

Another important aspect of our study was to examine the override rates by allergen-prescribed medication match. Similar to other reports [6], only about one-tenth of alerts were triggered by an exact match while the rest were triggered by probable matches of medication-allergy group. Not surprisingly, providers were significantly more likely to override probable matches, especially for non-immune mediated and non-life threatening reactions. Definite match allergy-medication combinations, especially those with life threatening reactions, were overridden the least, in about 70% of the cases. Still, the override rate was notably high and there is a need for further investigation, potentially using qualitative methods to better understand providers' reasons for overriding even the most alarming DAI alerts.

Finally, examining providers' reasons to override DAI alerts, we found that about one-third of the alerts were triggered for medications that patients have either previously tolerated or had no allergy to. Clearly, CPOE user interfaces must evolve to learn or negate medications patients have tolerated or had no allergy to in the past so providers are not inundated with alerts to previously tolerated medications [2]. Of note, we did not identify changes in DAI alert override patterns when we repeated the analysis without those categories. Another concerning trend was the lack of override reasons for about 45% of alerts. This finding raises several liability and ethical questions and warrants careful re-examination of the user interface to improve the capture and storage of DAI alert override reasons. Training of providers on importance of information captured by electronic health records might be one of the keys to this issue [20,21]. Finally, we also conducted some preliminary interviews with the providers to better understand why alerts are frequently overridden. This anecdotal evidence suggests that inpatient providers are reluctant to remove or otherwise change patients' allergies as they believe that this is a primary care provider's responsibility.

Limitations

This study is not without limitations. First, the analysis was limited to two large academic hospitals in Boston. In addition, different CPOE systems have different approaches to DAI alerting and use different medication vocabularies for drug-allergy interaction. Those factors might decrease the generalizability of our findings. Finally, our classification of immune mediated and life threatening opioid reaction is an estimation based on the captured patient reactions rather than a conclusive statement based on allergy tests.

Conclusions

This study examined trends in drug allergy interaction alerts for the most common trigger – opioid medications. Over the past decade, we found an increasing rate of DAI alert overrides culminating in a 90% rate. Allergic reaction information in our database included a high proportion (over 40%) of non-immune opioid reactions (e.g. gastrointestinal upset). The DAI override rate did not seem to differ significantly for immune mediated and life threatening reactions. The only factor contributing to the decrease in overrides was the nature of the DAI alert: definite allergy-medication matches were overridden less frequently compared to probable matches. Finally, about one-third of the alert override reasons pointed to the irrelevant alerts (i.e. "Patient has tolerated the medication before") and a significant proportion of DAI overrides were missing the override reason. Those findings warrant further investigation into providers' reasons for overriding alerts. It is also clear that user interfaces and drug-alerting algorithms should evolve to enable less interruptive and more accurate alerts to decrease alert fatigue.

Aknowledgments

This study is a part of a larger research project titled "Encoding and Processing Patient Allergy Information in EHRs" aimed at creating natural language processing tool for allergy information (R01HS022728-01).

References

[1] Office of National Coordinator for Health IT. Meaningful Use: Computerized Physician Order Entry (CPOE) for Medication Orders. 2104; Available at: http://www.healthit.gov/providers-professionals/achieve-meaningful-use/core-measures/cpoe-meaningful-use.

[2] Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc 2007; 14(1): 29-40.

[3] Nanji KC, Slight SP, Seger DL, Cho I, Fiskio JM, Redden LM, Volk LA, Bates DW. Overrides of medication-related clinical decision support alerts in outpatients. J Am Med Inform Assoc 2014; 21(3): 487-491.

[4] Smithburger PL, Buckley MS, Bejian S, Burenheide K, Kane-Gill SL. A critical evaluation of clinical decision support for the detection of drug-drug interactions. Expert Opin Drug Saf 2011; 10(6): 871-882.

[5] Abookire SA, Teich JM, Sandige H, Paterno MD, Martin MT, Kuperman GJ, Bates DW. Improving allergy alerting in a computerized physician order entry system. Proc AMIA Symp 2000: 2-6.

[6] Hsieh TC, Kuperman GJ, Jaggi T, Hojnowski-Diaz P, Fiskio J, Williams DH, Bates DW, Gandhi TK. Characteristics and consequences of drug allergy alert overrides in a computerized physician order entry system. J Am Med Inform Assoc 2004; 11(6): 482-491.

[7] Ariosto D. Factors Contributing to CPOE Opiate Allergy Alert Overrides. AMIA Annu Symp Proc 2014: 256-265.

[8] Joint Task Force on Practice Parameters- American Academy of Allergy, Asthma and Immunology. Drug allergy: an updated practice parameter. Ann Allergy Asthma Immunol 2010; 105(4): 259-273.

[9] DeDea L. Prescribing opioids safely in patients with an opiate allergy. JAAPA 2012; 25(1): 17.

[10] van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006; 13(2): 138-147.

[11] Kuperman GJ, Marston E, Paterno M, Rogala J, Plaks N, Hanson C, Blumenfeld B, Bates DW. Creating an enterprise-wide allergy repository at Partners HealthCare System. AMIA Annu Symp Proc 2003: 376-380.

[12] Microsoft™. Microsoft SQL Server Management Studio. 2008.

[13] StataCorp. STATA. v. 2010.

[14] Daubresse M, Chang HY, Yu Y, Viswanathan S, Shah ND, Stafford RS, Kruszewski SP, Alexander GC. Ambulatory diagnosis and treatment of nonmalignant pain in the United States, 2000-2010. Med Care 2013; 51(10): 870-878.

[15] Mazer-Amirshahi M, Mullins PM, Rasooly I, van den Anker J, Pines JM. Rising opioid prescribing in adult U.S. emergency department visits: 2001-2010. Acad Emerg Med 2014; 21(3): 236-243.

[16] Jena AB, Goldman D, Weaver L, Karaca-Mandic P. Opioid prescribing by multiple providers in Medicare: retrospective observational study of insurance claims. BMJ 2014; 348: g1393.

[17] Manchikanti L, Helm S, Fellows B, Janata JW, Pampati V, Grider JS, Boswell MV. Opioid epidemic in the United States. Pain Physician 2012; 15(3 Suppl): ES9-38.

[18] Khajouei R, Jaspers MW. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: a systematic review. Methods Inf Med 2010; 49(1): 3-19.

[19] Kuperman GJ, Gandhi TK, Bates DW. Effective drug-allergy checking: methodological and operational issues. J Biomed Inform 2003; 36(1-2): 70-79.

[20] Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2006; 13(5): 547-556.

[21] Agency for Healthcare Research and Quality. Computerized Provider Order Entry - Inpatient Best Practices. 2013: Available at: http://healthit.ahrq.gov/ahrq-funded-projects/emerging-lessons/computerized-provider-order-entry-inpatient.

Address for correspondence

Maxim Topaz PhD, RN, MA, Harvard Medical School and Brigham Womens Health Hosptial, 93 Worcester st. Wellsley Gateway, suite 2030I, Wellsley, MA 02481; email: mtopaz80@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-247

Usability Testing of PROCEnf-USP: A Clinical Decision Support System

Heloisa Helena Ciqueto Peres[a], Diná de Almeida Lopes Monteiro da Cruz[a], Michelle Tellez[b], Rita de Cassia Gengo e Silva[a], Diley Cardoso Franco Ortiz[c], Regina Celia dos Santos Diogo[a]

[a]*School of Nursing, University of Sao Paulo, Brazil*
[b]*Department of Nursing and Health Sciences, California State University East Bay, USA*
[c]*University Hospital, University of São Paulo, Brazil*

Abstract

Decision support systems (DSSs) are recognized as important tools, capable of processing high volumes of data and increasing productivity. The usability of these tools affects their effectiveness. By evaluating the interactions between registered nurses (RNs) and the DSSs, this study explores how they impact RN decision-making. This study analyzed 24 months (2011-2012) of data collected in Brazil in two units of a large, public, urban hospital in São Paulo that uses a nurse documentation system with an embedded DSS based on NANDA-I. Using mixed effects logistic regression, this study analyzed the agreement between RNs and a DSS when selecting nursing diagnoses. Results suggest that the agreement is mediated by characteristics of the RNs (education and experience) as well as units and year of encounter. Surprisingly, disagreement between RN and DSS when selecting defining characteristics (DC) had positive effects on the odds of agreement on diagnoses. Our results suggest that DSSs support nurses' clinical decision making, but the nurse's clinical judgment is the mediating factor. More research is necessary.

Keywords:

Nursing Information System; NANDA-I; Nursing Diagnoses, Classification System; Behavioral Model.

Introduction

The Institute of Medicine compiled several reports that support the use of technology to improve the quality and the safety of healthcare [1-4]. Computers are able to structure information and communication, becoming an intellectual tool. In this context, decision support systems (DSSs) are recognized as important tools, capable of processing high volumes of data and increasing productivity. Nurse focused DSSs collect, store, process, retrieve, display and transmit information in real time to assist registered nurses (RNs) with their work.

Through the nursing process nurses continuously assess, diagnose, plan, intervene, and evaluate their patients. All the while, nurses document their steps. PROCEnf-USP (Electronic Documentation System of Nursing Process) is a DSS – a product of a project funded by the National Council of Scientific and Technological Development (CNPq), a Brazilian agency of the Ministry of Science and Technology, and by the University of São Paulo (USP). The project has three aims: to develop computer systems, to install them, and to evaluate them. The PROCEnf-USP was designed by a group of faculty, students, and nurses from the School of Nursing, the Informatics Department, and the University Hospital (HU) of the USP, Brazil. The PROCEnf-USP was implemented in the surgical and clinical units at UH of the University of São Paulo in 2006 [5].

The system offers the RN a set of questionnaires to guide the RN documentation when performing patient assessments. The questionnaire has drop down menus with answers. Using these answers, the system computes probabilities for defining characteristics (DC) and presents them to the nurse. The nurse responds by selecting the DCs deemed most important. Once again the system uses the nurse's input and computes a set of nursing diagnoses (Dx) that best fit the clinical scenario. In special instances, the responses at the assessment may trigger Dxs without triggering DCs. The algorithm is based on the International Nursing Diagnoses Classification (NANDA-I), [6] which is being embraced worldwide. The system therefore can influence the way nurses think about priorities and patient care [7].

Nurses prioritize a DC or Dx based on intrinsic nurse characteristics. The intrinsic characteristics of the nurse deemed as relevant contributors to the effectiveness and efficiency of an individual in the workplace are human capital characteristics, such as level of education and years of experience [3]. These assumptions predict that nurses who are more educated and with more years of experience make more accurate nursing diagnoses.

Using the Donabedian's [8] quality framework, quality is analyzed through a relationship between structures, processes, and outcomes. Kelley, Brandon, and Docherty [7] apply this framework to the study of the effect of electronic nursing documentation on the quality of patient outcomes (Figure 1). Patient health outcomes are thought to be derived from high quality processes (what is done by the nurse to the patient) and high quality structures (human capital, setting, and equipment used when caring for patients).

With that in mind, this study aims to test the usability of the PROCEnf-USP in clinical practice by evaluating the agreement between the RN and the DSS when selecting Dx. The overarching goal is to better understand how the system is used by the RNs to promote better RN decision-making, and consequently improve patient outcomes.

According to this model, the interaction between the RN and the DSS is a key process (Figure 1). Structural aspects affecting this interaction are the characteristics of the nurse and the environment where she practices. The model purports that although the NANDA-I DSS system may be highly accurate, it is only capable of impacting patient health outcomes through interventions mediated by the nurse [7]. When applying the model to the PROCEnf-USP, it suggests that the DSS calculates the most complete list of DC and Dx;

therefore presenting information that may not have been obvious to the nurse. The nurse then selects the DC and Dx deemed most important and continues on with the other parts of the nursing process.

Figure 1 – Donabedian according Kelley et al [8,9].

Materials and Methods

This study is a secondary data analysis of existing observational nurse documentation data collected at a large, public, urban hospital in São Paulo, Brazil between January 1, 2011 and December 31, 2012. The unit of analysis is the admission encounter (N=337, 056).

Using all 24 months, this study analyzed the effects of several RN characteristics on the agreement between the RN and the DSS when selecting a nursing diagnosis, given the assessment data documented by the RN. A generalized linear mixed effects model for binomial family (mixed effects logistic regression) was conducted to assess the odds of agreement given the variables in the model [10,11] (Table 1).

The outcome (dependent) variable was represented by a dichotomous variable created by combining the four possible options. Agreement = yes, when the diagnosis was calculated by the DSS and selected by the nurse OR when the diagnosis was not calculated by the DSS and not selected by the nurse. Disagreement = no, when the diagnosis was calculated by the DSS but not selected by the nurse OR when the diagnosis was not calculated by the DSS but it was selected by the RN.

There were 18 nurses who documented using PROCEnf-USP between 2011 and 2012. RN variables included in the model were: age, level of education, years working at the HU, experience with research, and hours of continuing education in the past year. In addition, type of clinic (medical or surgical) and patient assessment data were included in the model.

Patient data were represented in terms of defining characteristics (DC). Selection of key DCs is the gold standard for nursing diagnosis selection accuracy [12]. Dummy variables representing four combinations of agreement between the DSS and the nurse in the selection of key DCs were created. They represented (a) when the DC was calculated by the DSS and selected by the nurse (62.71%), (b) when the DC was calculated by the DSS but not selected by the nurse (12.68%), (c) when the DC was not calculated by DSS but it was selected by the nurse (23.05%), and (d) when the DC was not calculated by the DSS and was not selected by the nurse (1.57%).

Results

Results from the analysis of our model by year of encounter 2011 (N=166,061) and 2012 (N=170,995), as well as the analysis of both years combined suggest that agreement between the DSS and the nurses regarding the selection of a nursing diagnosis was statistically significant affected by the year of the encounter, the number of hours of continuing education a nurse received, the unit in which the nurse worked, and the nurse's decision-making regarding the selection of DC (Table 2). Agreement occurred in 73.48% (N=247,656) of the time.

The analysis of all 24 month combined is presented in Table 2. The odds of agreement between the DSS and RN when selecting a priority Dx decreased by 7.2% (IC 95%; 0.927-0.929) between 2011 and 2012. The odds increased by 1.3% (IC 95%; 1.001-1.026) for each hour of continuing education the nurse received in the past 12 months, and by 12.8% (IC 95%; 1.071-1.188) if the RN was working at surgical unit compared with working at medical unit.

The odds of agreement between the DSS and the RN when selecting a Dx was also affected by the agreement between the RN and the PROCEnf-USP when selecting DCs. Compared with agreement characterized as 'DC calculated by DSS and selected by the RN,' 'DC calculated by the DSS but not selected by the RN' increased the odds of agreement in the selection of the Dx 37.4 times (IC 95%; 34.56-40.58), and 'DC not calculated by the DSS but selected by the RN' increased the agreement by a third or 36.8 % (IC 95%; 1.33-1.40).

In contrast, compared with agreement characterized as 'DC calculated by DSS and selected by the RN', 'DC not calculated by the DSS and not selected by the RN' reduced the odds of agreement when selecting a priority nursing diagnosis by 99.4% (IC 95%; 0.005-0.008).

Level of education (BSN vs. MS/PhD), participation in research (Yes/No), years of experience at the HU, and age were not significant, and therefore did not affect the odds of agreement between the RN and the DSS when selecting Dx (Table 2).

Table 1 – Variables included in the model

		N	%
Education	MSN/PhD	6	33.33
	BSN	12	66.67
Research	Yes	9	50.00
	No	9	50.00
Clinic	Surgical	10	55.56
	Medical	8	44.44
	Mean	**SD**	**Range**
Continuing education	36.11	24.04	1-79
Age	36.94	7.40	28-48
Experience HU	10.33	6.01	2-20

Table 2 – Mixed Model Effect Logistic Regression 2011-2012

Fixed effects	**B**	**SE**	**Odds Ratio**	**IC 95% Inf–Sup**
Age	-0.013	0.037	0.987	0.917-1.062
Education	0.313	0.287	1.368	0.779-2.402
Research	-0.276	0.225	0.759	0.488-1.179
Years at HU	0.056	0.048	1.057	0.963-1.161
Continuous edu.	0.013**	0.006	1.013	1.001-1.026
Surgical Unit	0.121**	0.026	1.128	1.071-1.188
DC calculated but not selected by the RN	3.623**	0.041	37.450	34.564-40.576

DC not calculated but selected by the RN	0.313**	0.014	1.368	1.331-1.405
DC not calculated and not selected by the RN	-5.070**	0.102	0.006	0.005-0.008
Year	-0.075**	0.001	0.928	0.927-0.929

** P<0.01

Discussion

The most important finding of this study is that the agreement between the decision support system (DSS) and the registered nurse (RN) when selecting a nursing diagnosis (Dx) is dependent on the agreement between RN and DSS when selecting defining characteristics (DC), but not in the way it was anticipated. We expected that the agreement in the selection of Dx was dependent on the "literal" agreement between the DSS and RN regarding the selection of defining characteristics (DC). However, we found that varying degrees of agreement regarding DC may provide agreement between the RN and DSS when selecting a Dx.

As described in the methods, patient data were represented in the model through DC. There are four possible combinations of agreement between the RN and DSS on the selection of DC. One combination can be characterized as being the true agreement (i.e. when the DSS calculates a DC and the RN selects it). Two combinations are in fact disagreements with a 50:50 split between the DSS and the RN and one combination ,a false agreement, neither calculates nor selects a DC.

These variations in agreement occur when the RN completes the nursing assessment phase of the encounter, but does not complete the documentation thoroughly. Since the documentation is the basis for the calculation of DCs, the system is bound by the information inputted. Therefore, when selecting or ignoring a DC, nurses are editing the data from the assessment phase.

For instance, in 23.05% of the admission cases between 2011 and 2012, the RN selected a DC that was not calculated by the PROCEnf-USP, but ultimately agreed with the system when selecting a Dx. In fact, in these instances there was an increase of 36.8% in agreement between DSS and the RN in the selection of Dx. This supports the idea that the RN is the mediator for nursing diagnostic decision-making. The RN may skip the step of entering the data, bypassing the NANDA-I algorithm, and going straight to the diagnosis.

Conversely, in 12.68% of the admission encounters between 2011 and 2012, the RNs did not select a DC calculated by PROCEnf-USP, but ultimately agreed with the system in the selection of a Dx. In fact, in these instances the odds of agreement increased by a large margin, 37 times. By ignoring a DC calculated by the DSS but agreeing with the DSS on the priority diagnosis, the nurse is suggesting that the DC calculated by the DSS is less important that it appears. It is possible that the RN is providing contextual information that was not captured in the documentation.

In 1.57% of the admission encounters between 2011 and 2012, the RN did not select and the DSS did not calculate a DC. In those instances there was a decrease in agreement by 99.4%. The algorithm is designed to allow the DSS to bypass the DC and directly suggest a Dx given dichotomous assessment questions. It is possible that in these instances, the DSS did not calculate a DC, but suggested a Dx. Since the DSS did not calculate the DC, the RN was less like to select it on their own, and consequently less likely to agree with the Dx calculated.

When the nurse adds information or ignores information, the RN's decisions are influenced by something other than the DC. The model suggests that RNs' human capital characteristics, such as experience working at that unit and hours of continuing education may have had an impact. Our results support that. For each additional hour of continuous education a nurse received in the past 12 month, there was an increase of 1.3% in the odds of agreement between the RN and DSS when selecting a nursing diagnosis. At University Hospital, nurses received a mean of 36.11 hours per year of continued education. Although some of the topics included informatics, nursing taxonomy, and nursing documentation, most topics were not specific for DSS usage. They provided or reinforced clinical knowledge, which is essential for clinical reasoning development [13].

The odds of agreement was 12.8% higher in the surgical unit than in the medical unit. Although this finding appears to be positive, it is of concern. Patient turnover is higher in the surgical unit than it is in the medical ward; the average length of stay is shorter, but patient profiles are very similar. Therefore RNs are required to work faster during the assessment. They have less time to be in contact with their patients, but are experiencing similar clinical conditions. It is possible that RNs in the surgical units are thinking that they know a priori what their patients' problems are. The RNs at the surgical unit may be collecting patterns of patient data, inputting them in the DSS, and then agreeing with the system. In 2012 the odds of agreement between the RN and the DSS decreased by 7.2% when compared to 2011. This may be a result of inexperience, given that in 2012 the medical unit had four new RNs who transferred from other units of HU [14].

Conclusion

The agreement between the RN and the DSS when selecting defining characteristics (DC) significantly affects the odds of agreement between the RN and the DSS when selecting nursing diagnoses (Dx), but not as anticipated. The system offers the structure for nurses to collect data and analyze information by offering opportunities for nurses to accept or reject DC and Dx. These interactions between the RN and DSS are mediated by the RN characteristics; such as years of experience and continuing education. These characteristics are proxies for clinical reasoning and clinical judgment. Systems such as PROCEnf-USP are especially helpful to students, new nurses, or expert nurses in a new environment. Future studies should compare the documentation of the novice RNs and students to those of more experienced RNs. These results along with future studies in this field will contribute to the refinement of the Dx classification based on real clinical environment evidences.

References

[1] Kohn L I, Corrigan J M, Donaldson M S. To Err is Human: Building a Safer Health Care System. Washington, DC: Institute of Medicine, 2000.
[2] Committee on Quality of Health Care in America. Crossing the Quality Chasm: A new health system for the 21st Century. Washington, DC: Institute of Medicine, 2001.
[3] Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing. The future of nursing:

Focus on education. In National Academies of Science (Ed.). Washington DC: Institute of Medicine, 2010.
[4] Smith M, Saunders R, Stuckhardt L, McGinnis J M. Best care at Lower cost: The path to continuously learning healthcare in America. Institute of Medicine. Washington DC: Institute of Medicine, 2012.
[5] Peres HHC, Cruz DALM, Lima AFC, et al. Desenvolvimento de Sistema Eletrônico de Documentação Clínica de Enfermagem estruturado em diagnósticos, resultados e intervenções. Rev. esc. enferm. USP. 2009;43(2):1149-55
[6] Diagnósticos de enfermagem da NANDA: definições e classificações 2007-2008; tradução Regina Machado Garcez. Porto Alegre (RS): Artmed, 2007.
[7] Kelley, T. F., Brandon, D. H., & Docherty, S. L. (2011). Electronic nursing documentation as a strategy to improve quality of patient care. *Journal of Nursing Scholarship, 42*(2), 154-162.
[8] Donabedian A. The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press; 1980.
[9] Scheffler RM, Bruckner TA, Spetz J. The labour market for human resources for health in low- and middle-income countries. Human Resources for Health Observer. 2012(110
[10] AIC: H. Akaike (1974) A new look at the statistical model identication. IEEE Transactions on Automatic Control, 19(6):716-723.
[11] BIC: G. Schwarz (1978) Estimating the dimension of a model. The annals of statistics, 6(2):461-464.
[12] Martins, TYC. Avaliação da acurácia de diagnósticos de enfermagem em sistema de documentação eletrônica [dissertação]. São Paulo: Escola de Enfermagem, Universidade de São Paulo; 2014.
[13] Lunney, M. Critical thinking and accuracy of nurses' diagnoses. Part II: application of cognitive skills and guidelines for self-development. Rev. esc. enferm. USP [online]. 2003, vol.37, n.3.
[14] Benner P. From Novice to Expert. The american Journal of Nursing. Lippincott Williams & Wilkins. Vol. 82, Nº 3 (Mar., 1982) pp 402-407 Disponível em: http://www.jstor.org/stable/3462928. Acessado em 01 de outubro de 2014.

Address for correspondence

Regina Célia dos Santos Diogo, RN
PhD Candidate, School of Nursing, University of Sao Paulo
Av Dr Eneas de Carvalho Aguiar, 419 – 1º andar
Cerqueira Cesar – São Paulo CEP 05403-000
Phone: + 55 (11) 30617552/ + 55 (11) 996667974
E-mail: regina_diogo@terra.com.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-251

Patient Safety in Critical Care Unit: Development of a Nursing Quality Indicator System

Camila S.P. Lima[a], Sayonara F.F. Barbosa[a]

[a] *Federal University of Santa Catarina (UFSC), Brazil.*

Abstract

This is a methodological study and technological production that aims to describe the development of a computerized system of nursing care quality indicators for the Intensive Care Unit. The study population consisted of a systems analyst and fifteen critical care nurses. For the development of the system we adopted some of the best practices of the Unified Process methodology using the Unified Modeling Language and the programming language Java Enterprise Edition 7. The system consists of an access menu with the following functions: Home (presents general information), New Record (records the indicator), Record (record search), Census (add information and indicators of the patient), Report (generates report of the indicators) and Annex (accesses the Braden Scale). This information system allows for measurement of the quality of nursing care and to evaluate patient safety in intensive care unit by monitoring quality indicators in nursing.

Keywords:
Nursing Informatics; Quality Indicators; Intensive Care Unit.

Introduction

Applications of information technology in nursing aim for knowledge systematization to quantify management and nursing care [1]. In the area of patient safety, a Health Information System can be developed in order to reduce the occurrence of adverse events by improving the quality of care practice using quality indicators, which are screening tools for the purpose of identifying potential areas concerning clinical care quality. Quality indicators also reflect the quality of care in hospitals and should bring to light what is usually invisible in nursing, rather than what is visible only when absent [2,3]. Therefore, it is essential to control indicators to improve clinical care and to achieve the desired results. Among the various health care scenarios, the Intensive Care Unit (ICU) stands out as an environment where the focus on patient safety should be strongly present [4]. Here, aspects such as complexity of care, the severity of the patient's disease, large amount of drugs and invasive devices, and the multiple procedures performed, make the ICU environment quite vulnerable to adverse events. It raises the question: how to develop a system to register and analyze quality indicators specific for critical care nursing? It is believed that the use of computerized tools can contribute to obtaining desirable results related to health practices, providing patient safety and decreasing the occurrence of adverse events. This study aims to describe the development of a computerized Quality Indicators Nursing System (SIQenf) for an Adult Intensive Care Unit.

Methods

This is a methodological study and technological production designed for the ICU of a university hospital in southern Brazil. It had the participation of a systems analyst and fifteen nurses in intensive care. The inclusion criteria of the study participants were, a) to be a critical care nurse to validate quality indicators and their related factors and also to use the computerized system; b) Systems Analyst - graduate in Computer Science or Information Systems. The period of data collection (system development period) was from February to April 2014. Data were entered into the system by the nurses that accepted to participate in the study. They were registered and received a username and password, maintaining the anonymity of the participants. For the development of the system, some of the best practices of the Unified Process methodology were used [5]. This process is divided into four phases: inception, elaboration, construction and transition. The inception phase identified the requirements that were described in terms of use cases, and presented which features and functions are desirable for each class of users [6, 7]. Use cases can be represented by Use Case Diagrams (Figure 1).

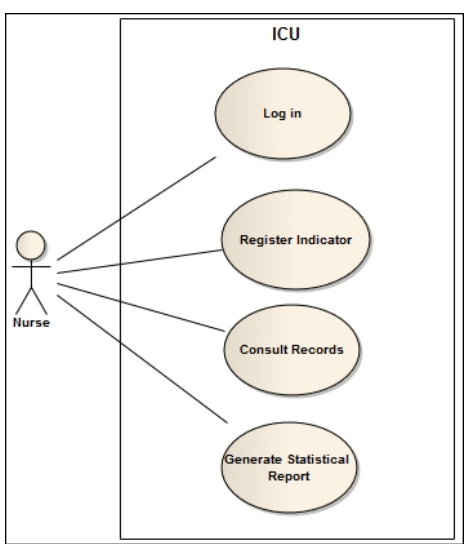


Figure 1- SIQenf use case diagram

The content validation of quality indicators in nursing was performed at this phase and also its related factors (causes) with the ICU nurses who participated in the study. Eight indicators were selected: 1) Fall Incidence; 2)Incidence of non-planned tracheal extubation; 3)Loss of Oro / Nasogastroenteral tube; 4) Incidence of Pressure Ulcer; 5) Medication error; 6) Near miss related to Drug Administration; 7) Phlebitis Incidence and; 8) Lost of Central

Venous Catheter. After the definition of the desired indicators, the participants also identified 73 related factors (causes) of the indicators. These indicators were selected because they are used in various health organizations and because research shows the incidence rates of them in intensive care areas [8-15].

In the elaboration phase, the system was modeled using the Unified Modeling Language (UML). In the construction phase the software components that will make each use case operational for the end users were developed. And in the last phase of transition, a training was carried out with all the nurses of the study, enabling them to use the computerized system. For the development of the system the object-oriented programming language Java Enterprise Edition 7 object (Java EE 7) was used. An Apache Tomcat server was used for the Java application and we also used the Java Server Faces framework (JSF) for the development of the web interface. As a Database Management System (DBMS), a relational open source MySQL version 5.6 was chosen. The study was approved by the Ethics Committee of the University where the study was conducted.

Results

The system was named SIQenf (Nursing Quality Indicator System) and was developed for use in an intensive care unit. The SIQenf aims to support the healthcare practice, allowing the measurement of quality of patient care through the use of quality indicators, and to assist in management and research.

System Functional Characteristics

Login screen

Each nurse that participated in the study received an username and password. The system is accessed through the Internet address: giate.ufsc.br/SIQenf.

Home Screen

The System Home screen consists of a menu at the top of the screen that has the following functions: Home, New Record, Record, Census Report, Annex and Exit.

New Entry screen

The SIQenf allows nurses to record the quality in nursing indicators (Figure 2) and its related factors (causes), selected from among eight indicators: (1) Fall Incidence; (2) Incidence of unplanned extubation of endotracheal cannula; (3) Incidence of unplanned Oro/Nasogastroenteral tube removal; (4) Incidence of Ulcer Pressure; (5) Medication error; (6) Near miss Drug Administration; (7) Incidence of Phlebitis and; (8) Unplanned retrieval of Venous Catheter Central.

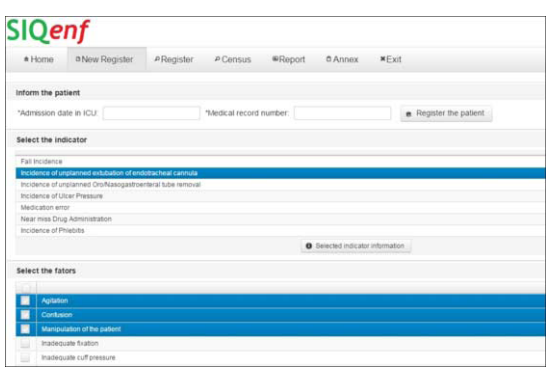


Figure 2 – New registration screen

By accessing this screen, nurses can view the indicators recorded on the current date. The "Filter Time" function reports the desired period and search the recorded indicators.

Census screen

To calculate the indicator it is necessary to insert a new census, i.e., the daily number of hospitalized patients, the number of intubated patients, the number of patients with oro/nasogastroenteral tubes, the number of patients with a risk for ulcer pressure, the number of patients with peripheral venous access and the number of patients with central venous catheters.

Report Screen

On this screen it is possible to generate the general report indicators (Figure 3), which is available in Portable Document Format (PDF), where the value of all indicators is calculated using their respective formulas, and provides the index (%) and graph of each indicator in the selected period. On this screen it is also possible to select individual report indicators, and this report will show the definition of the indicator, its formula, and its value in the period.

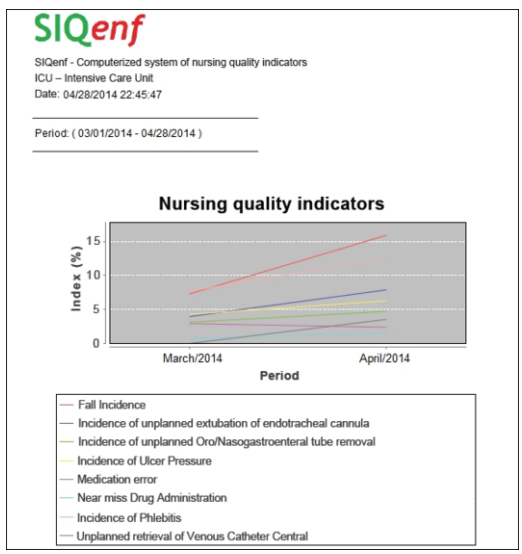


Figure 3 – General Indicators Report

Annex Screen

In this screen it is possible to download the Braden Scale, which must be used daily in the ICU to assess the risk of pressure ulcer development. This scale has a maximum score of 23 and a minimum of 6 points. It evaluates the following aspects: sensory perception, moisture, physical activity, mobility, nutrition, and friction and shear [16]. In the system a Braden score $\leq$ 16 was adopted for risk of development of ulcer pressure.

Discussion

SIQenf was developed in order to facilitate the measurement of quality and evaluation of patient safety in intensive care, since the indicators allow us to measure the level of health actions.

The idea of developing a computerized system of indicators rather than a manual system emerged in order to facilitate the management of nursing care, since the computer facilitates the integration of data, information and knowledge [17]. It

represents a resource to facilitate nursing registration, making it faster and more accurate, providing nursing and other professionals updated and reliable information, and has the potential to transform the quality of care and establish links between nursing care and results for the patients [18-19]. There are a significant number of studies, predominantly international, seeking information technology solutions to improve patient safety and to bring this issue to the area of intensive care. Research shows there are a high number of errors and adverse events in the ICU, and emphasizes the importance of assessing nursing care in this scenario [20-22].

Studies have assessed the quality of care and patient safety in the ICU through nursing quality indicators, where the incidence of medication errors, loss of nasogastroenteral probe, loss of central venous catheter and incidence of pressure ulcers were highlighted, however these surveys were developed using manual systems for the analysis of indicators [23-25]. In Brazil, the number of studies using computerized systems for detecting adverse events and evaluation indicators is still low. As an example, one study described the steps of building a computerized system for managing nursing care indicators of the Hospital São Paulo, using the following indicators: incidence of unplanned extubation, hypothermia, skin lesions, loss of urinary catheter, loss of central venous catheters, peripheral venous catheters loss, loss of drain tube, loss of the gastrointestinal tract tube, patient falls and pressure ulcers [26].

It can be observed that most information systems in the patient safety area focus on reporting events and to improve the safety of the patient, it should go beyond the registration report and use, and therefore identify data security risks, develop interventions to mitigate these risks, and evaluate whether interventions have reduced or prevented the damage [27].

The use of SIQenf can make it possible to detect adverse events, to record and analyze the quality indicators, as well as the major factors that are involved and that contributed to their occurrence. Thus, in the future preventive measures may be applied to minimize or even reduce the incidence of these indicators in the intensive care unit, through training, continuing education and safe interventions based on scientific evidence.

Conclusion

The use of information technology by nurses has been an increasingly constant practice, in clinical environments, management, teaching, or research. This integration of information technology and nursing facilitates the management of health information and contributes to the quality of patient care. On this basis, this paper describes the development of a computerized system of Nursing Quality Indicators (SIQenf) for the ICU, which supports the nursing care practice, which can measure the quality of care and evaluate patient safety through the detection of adverse events and monitoring quality indicators in nursing. It is believed, however, that SIQenf is more than a tool for reporting events, as it can contribute to the promotion of quality and strengthening patient safety in the ICU. It records and calculates the indicators and their related factors (causes), making it possible to understand the main problems related to patient safety in intensive care and make interventions to decrease these problems, as well as help in the search for solutions and excellence in nursing care. In addition, it is also possible through the reports to identify the main incident indicators according to the period and to compare them with other ICU indicators. For future work, this system will undergo an evaluation period of ergonomic criteria and usability, and will subsequently be implemented in the ICU for which it was developed.

Acknowledgments

To the National Council for the Improvement of Higher Education (CAPES) for providing a Master's scholarship to Camila Lima.

References

[1] Cavalcante RB, Ferreira MN, Silva LTC, Silva PC. Experiências de informatização em enfermagem no Brasil: um estudo bibliográfico. J. Health Inform 2011: 3(3): 130-4.

[2] Agency for Healthcare Research and Quality. At A Glance. Rockville, MD: Agency for Healthcare Research and Quality 2011 [accessed on September 1, 2014]. Available from: http://www.qualityindicators.ahrq.gov/

[3] Curley M, Kapito R. Quality & Safety - How do we measure it. Nurse-Sensitive Quality Outcomes 2012 [accessed on November 20, 2014]. Available from: http://www.criticalcarecanada.com/presentations/2012/nurse-sensitive_quality_outcomes.pdf

[4] Barra DCC, Sasso GTM, Baccin CRA. Sistemas de alerta em um processo de enfermagem informatizado para Unidades de Terapia Intensiva. Rev Esc Enferm USP 2014: 48 (1):127-34.

[5] Jacobson I, Booch G, Rumbaugh J. The Unified Software development Process. Mass:Addison-Wesley, 1999.

[6] Hannah KJ, Ball M, Edwards MJA. Introdução à Informática em Enfermagem. 3th ed. Porto Alegre: Artmed, 2009.

[7] Pressman RS. Engenharia de Software. 6th ed. Porto Alegre: MCGrawHill, 2010.

[8] American Nurses Association. National Database of Nursing Quality Indicators [accessed on September 16, 2014]. Available from: http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume122007/No3Sept07/NursingQualityIndicators.html

[9] Ray B, Samaddar DP, Todi SK, Ramakrishnan N, John G, Ramasubban S. Quality indicators for ICU: ISCCM guidelines for ICUs in India. Indian J Crit Care Med 2009:13:173-206.

[10] Joint Commission International. Implementation Guide for the NQF Endorsed Nursing-Sensitive Care Measure Set 2010 [accessed on November 20, 2014]. Available from: http://www.jointcommission.org/assets/1/6/NSC%20Manual.pdf

[11] Australian Council on Healthcare Standards. Intensive Care Clinical Indicators User Manual 2011[accessed on December 10, 2014].Available from: http://www.achs.org.au/ pirt.

[12] Sociedade Spañola de Medicina Intensiva, Crítica y Unidades Coronarias. Quality Indicators in Critically Ill Patients 2011[accessed on December 05, 2014]. Available from: http://www.semicyuc.org/sites/default/files/quality_indicators_update_2011.pdf.

[13] Tronchin DMR et al (2012). Manual de indicadores de enfermagem NAGEH / Compromisso com a Qualidade Hospitalar (CQH) 2 ed. São Paulo(SP): APM/ CREMESP 2012 [accessed on November 18, 2014]. Available: http://www.cqh.org.br/ portal/pag/doc.php?p_ndoc=125.

[14] Flaatten H. The present use of quality indicators in the intensive care unit. Acta Anaesthesiol Scand 2012: 56:1078–1083.

[15] Scottish Intensive Care Society. Quality Indicators for Critical Care in Scotland 2012 [accessed on September 15, 2014]. Available from: http://www.scottishintensivecare.org.uk/uploads/2014-01-12-22-37-12-QualityIndicators2012pdf-62756.pdf.

[16] Araujo CRD, Lucena STM, Santos IBC, Soares MJGO. A enfermagem e a utilização da Escala de Braden em úlcera por pressão. Rev Enferm UERJ 2010:18(3):359-64.

[17] Barbosa SFF, Sasso GTM. Informática na pesquisa em enfermagem. Revista Eletrônica de Enfermagem 2009: 11:724-31.

[18] Dykes PC, Collins SA. Building Linkages between Nursing Care and Improved Patient Outcomes: The Role of Health Information Technology. OJIN: The Online Journal of Issues in Nursing 2013:18(3).

[19] Veríssimo RCSS, Marin HF. Protótipo de sistema de documentação em enfermagem no puerpério. Acta Paul Enferm 2013:26(2):108-15.

[20] Silva EWNL, Araujo RA, Oliveira EC, Falcão VTFL. Aplicabilidade do protocolo de prevenção de úlcera de pressão em unidade de terapia intensiva. Rev Bras Ter Intensiva 2010:22(2):175-185.

[21] De Groot RI, Dekkers OM, Herold IH, de Jonge E, Arbous MS. Risk factors and outcomes after unplanned extubations on the ICU: a case-control study. Crit Care 2011:15(19):1-9.

[22] Garrouste-Orgeas M, Philippart F, Bruel C, Max A, Lau N, Misset B.Overview of medical errors and adverse events. Ann Intensive Care 2012:2(1):2.

[23] Matos LS, Duarte NLV, Minetto RC. Incidência e prevalência de úlcera por pressão no CTI de um Hospital Público do DF. Rev. Eletr. Enf 2010:12(4):719-26.

[24] Beccaria RLM, Pereira LAM, Contrin M, Lobo SMA, Trajano DHL. Eventos adversos na assistência de enfermagem em uma unidade de terapia intensiva. Rev Bras Ter Intensiva 2009:21(3):276-82.

[25] Nascimento CCP, Toffoletto MC, Gonçalves LA, Freitas WG, Padilha, KG. Indicadores de resultados da assistência: análise dos eventos adversos durante a internação hospitalar. Rev. Latino-Am. Enfermagem 2008:16(4):746-751.

[26] Labbadia LL, D'Innocenzo, M, Flogliano RRF et al. Sistema Informatizado para Gerenciamento de Indicadores da Assistência de Enfermagem do Hospital São Paulo. Rev Esc Enferm USP 2011:45(4):1013-7.

[27] Pronovost PJ, Morlock LL, Sexton JB et al. Improving the value of patient safety reporting systems. In: Henriksen K, Battles JB, Keyes MA, et al., eds. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment). Rockville (MD): Agency for Healthcare Research and Quality [accessed on November 20, 2014]. Available from: http://www.ncbi.nlm.nih.gov/books/ NBK43621

Address for correspondence

Camila Santos Pires Lima

Rua João Motta Espezim, 859

Saco dos Limões, Florianopolis, Santa Catarina, Brazil.

ZIP CODE: 88045401.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-255

Pre-Implementation Study of a Nursing e-Chart: How Nurses Use Their Time

Maria B. Schachner[a], Francisco J. Recondo[b], Janine A. Sommer[b], Zulma A. González[b], Gabriela M. García[b], Daniel R. Luna[b], Sonia E. Benítez[b]

[a] *Nurse Department of Hospital Italiano de Buenos Aires*
[b] *Department of Health Informatics de Buenos Aires*

Abstract

In clinical practice, nurses perform different activities that exceed direct care of patients, and influence workload and time administration among different tasks. When implementing changes in an electronic nursing record, it is important to measure how it affects the time committed to documentation. The objective of this study was to determine the time dedicated to different activities, including those related to electronic documentation prior to implementing a redesigned nurse chart in an Electronic Health Record at the Hospital Italiano de Buenos Aires. An observational work sampling study was performed. Nursing activities observed were categorized as direct care, indirect care, support, non-patient related, and personal activities. During the study, 74 nurses were observed and 2,418 observations were made in the Intensive Care Unit (32.22%), the Intermediate Care Unit (29.57%), and the General Care Unit (38.21%). Nurses' activities included 37.40% of direct care, 41.18% of indirect care, 0.43% support tasks, 11.14% non-related to patient tasks, and 9.77% personal activities. The results allow for the estimation of the impact of a nursing e-chart on nurses' activities, workflow and patient care.

Keywords:

Work sampling study; Time and motion study; Non-direct care; Indirect care; Nursing workload; EHR use; EHR documentation.

Introduction

The shortage of nurses is a growing problem, and it is estimated that in the next few years the problem will intensify [1]. The existence of nursing activities and factors that are not related with direct patient care is known. Several authors have pointed out that the time devoted to these activities impact the nurses' workload [2–4]. In addition, according to Myny et al. [5], there are still doubts regarding which of these indirect patient care activities are perceived by nurses as part of their work responsibilities. Information technology could be leveraged to solve this problem through multiple implementations that take into account specific nurses' needs. There is no consensus in the literature on how the implementation of an Electronic Health Record (EHR) system may affect the efficient use of time. Several studies show a reduction in nursing documentation time [2,4], while others show that the time gained due to new functionalities in the EHR is lost. According to nurses, informed implementation of the new EHR system reduces communication with physicians, generating mistrust on certain prescriptions, which requires constant re-confirmation, leading to an increase in associated documentation time [6]. Consequently, this results in a lack of nursing staff, endangering work efficiency and generating a rise in hospital costs [7]. Given the impact on time standards of nursing care and human resources management, data regarding nursing time dedicated to different activities is crucial [8]. This information is necessary to evaluate changes in nursing practice after the implementation of a new system, and is decisive in evaluating its efficacy [9]. In order to measure both work and time, different techniques exist to generate standards, "time and motion studies", and "work sampling". The selection of these techniques will depend on the precision and nature of the object in study. These types of studies have been done to describe and evaluate the impact of health care professional activities, including physicians [10], pharmacists [11,12], and nursing professionals [13]. In the context of an implementation of a redesigned electronic nursing record at Hospital Italiano de Buenos Aires (HIBA), the objective of this article is to determine the time needed for nurses to do specific activities, including those related to electronic documentation prior to implementation.

Materials and Methods

Setting

HIBA is an academic tertiary level hospital founded in 1853. It belongs to a nonprofit healthcare network including 25 ambulatory clinics and 150 outpatient offices in Buenos Aires, Argentina. The infrastructure includes 750 beds (200 for critical care), 800 home care beds, and 41 operating rooms. There are 2,800 physicians, 2,800 non-physician healthcare professionals, and 1,900 individuals in administrative services and management. During 2013-2014 there were approximately 45,000 discharges, 45,000 surgical procedures (50% of which were ambulatory), and 3 million visits. Since 1998, HIBA has gradually developed and implemented an "in house" Health Information System (HIS) that handles medical and administrative information from point of capture to analysis. The HIS includes a single, modular, problem-oriented, and patient-centered electronic health record (EHR). The EHR, named ITALICA, allows for the recording of patient care at different levels (outpatient, inpatient, emergency, and home care). ITALICA also enables complementary studies, drug prescriptions, and results display that includes the storage and transmission pictures system (PACS - Picture archiving and communication system). In recent years, the nurse chart evolved from paper to the computerized system. The first phase was digitized paper

documents, but in 2010 the first version of the electronic nursing record was embedded in the EHR, consisting of sections of structured data entry, medication administration, fluid balance, vital signs and a free-text area where nurses record narrative observations. Two years later, the system was updated to a version organized by sections according to nursing process, including planning of nursing interventions (but not nursing diagnosis). Currently, the electronic nurse chart is structured in four sections independent of each other: Assessment, Planning, Implementation, and Evaluation. Different sections allow nurses to record patient care following the logic of nursing processes, or skip to any section without completing the others.

For the pilot, we chose three representative sectors: the Adult Intensive Care Unit (ICU), the Adult Intermediate Care Unit (IMCU), and the Adult General Care Unit (GCU). The ICU has 38 beds allocated in three areas, according to the severity and therapeutic requirement of the patient. The nurse:patient ratio is 1:2, and there are 108 nurses allocated to the different shifts. The IMCU has 28 beds, the nurse:patient ratio is 1:3, and the total number of nurses is 58. The GCU has 44 beds; it is a medical-surgical unit, with a nurse:patient ratio of 1:8, and 32 nurses are distributed across the different shifts.

Design

This study was an observational descriptive work sampling. Data collection was made through observations, and work samples were made following the steps proposed in the literature [7,14]. The task consisted of observing a work sample of nurses, and describing the activities performed as well as determining the time cost of those activities.

After the approval of the project, the task categories and the specifics activities were determined. After a literature review [9,15–17], the categories were agreed on with the Chief of Nursing Department (CND). The categories chosen are direct care, indirect care, support activities, non-patient related activities, and personal activities.

***Direct care*:** The activities near a patient's bed, including admission, anamnesis (medical history), comfort, emotional support, and education, among others.

***Indirect care*:** Documenting on paper or EHR, information exchange on handoffs, pharmaceutical preparations, and supplies for procedures typify this category.

Support activities: Interdepartmental activities, scheduled training, and coaching other nurses are examples of this kind of activity.

***Non-patient related activities (NPRA)*:** Activities associated with equipment search, arranging the unit, performing claims to suppliers, use of computers for non-patient related tasks, changing the patient to another bed within the same unit, and the "waiting time" to perform other tasks.

Personal activities: These activities include breakfast/lunch breaks, social interaction, and non-patient related conversations.

Three trained observers collected the data using a worksheet. The CND estimated the percentages of time spent on each category. The use of the EHR (or EHR documentation) was set to 15% [16]. The time duration analysis included mean, standard deviation (SD), minimum (Min), and maximum (Max).

Results

From November 17 to December 5, 2014, 2,418 observations were made. The ICU had 32.22% observations, the IMCU had 29.57%, and the GCU 38.21%. Seventy-four nurses were observed, 82% of whom were women, with mean age 37.56 years old (SD 9.46), and the mean seniority was 10.37 years (SD 8.95). Overall, the nurses' activities included 37.40% of direct care, 41.18% of indirect care, 0.43% support tasks, 11.14% non-patient related tasks, and 9.77% personal activities. The following tables show mean, standard deviation, minimum and maximum values for percentages from different observations, differentiated per sector and categories.

Table 1 shows the ICU data, where the direct care activities represent 34.55%, and EHR activities are 23.79% of the total.

Table 1 - Intensive care unit activities

%	Direct care	Indirect care	Support activities	NPRA	Personal activities	EHR use
Mean	34.55	44.51	0.00	13.14	7.80	23.79
SD	1.49	2.63	0.00	2.25	1.67	3.65
Min	32.69	41.94	0.00	9.93	5.77	19.88
Max	36.77	48.94	0.00	16.03	9.62	29.08
Estimated	35.00	30.00	20.00	5.00	10.00	15.00

The next table (Table 2) shows the IMCU data, where the direct care activities increase to 38%, and EHR activities are 18.26%.

Table 2- Intermediate care unit activities

%	Direct care	Indirect care	Support activities	NPRA	Personal activities	EHR use
Mean	38.61	36.79	0.14	10.69	13.77	18.26
SD	5.22	5.09	0.32	3.65	4.25	2.93
Min	34.44	30.82	0.00	6.34	8.89	15.23
Max	47.41	44.37	0.71	14.57	20.55	21.48
Estimated	40.00	30.00	15.00	10.00	5.00	15.00

The last table (Table 3) presents GCU unit data. Here, direct care activities are 39.05%, and EHR activities are 19.97%.

Table 3- General care unit activities

%	Direct care	Indirect care	Support activities	NPRA	Personal activities	EHR use
Mean	39.05	42.25	1.14	9.59	7.74	19.97
SD	7.51	2.68	1.06	3.32	3.32	2.17
Min	27.32	37.64	0.00	4.59	5.10	17.06
Max	46.63	44.33	2.06	12.89	13.40	23.12
Estimated	30.00	30.00	10.00	20.00	10.00	15.00

The following figures show the variations along the period of study. Figure 1 corresponds to the ICU. EHR documentation was the activity with the most fluctuation. At the beginning, 19.88% of the activities where related to the EHR, while at the end, these activities increased to 29.08%:

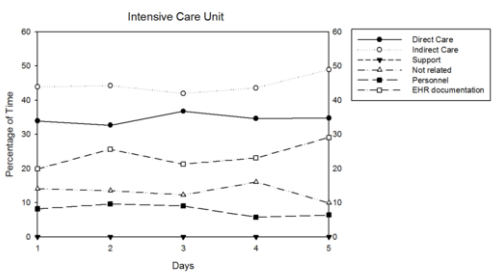


Figure 1- Observations at ICU during the study

In the IMCU (see Figure 2), all activities had large variations along the evaluation period, but none of them had big differences comparing the beginning and the end of the study.

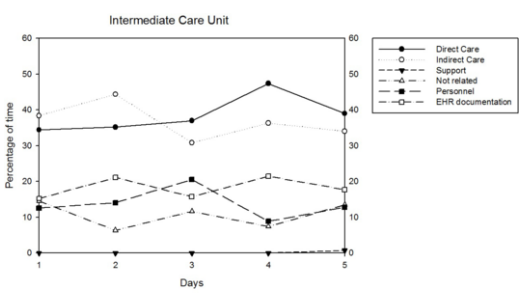


Figure 2- Observations at IMCU during the study

In the GCU, the direct care activities were found to be 27.32% at the beginning, and ended at 46.63%. Other categories did not have important disparities (See Figure 3).

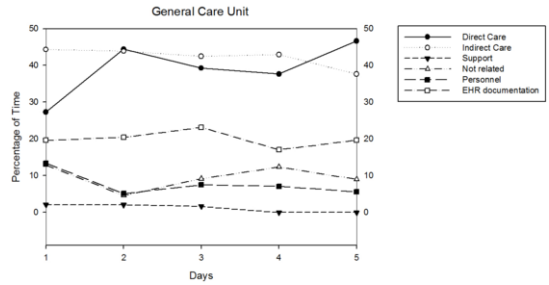


Figure 3 - Observations at GCU during the study

Discussion

In this pilot study, we evaluated how nurses use their time in the pre-implementation phase of redesigning a nursing e-chart. As part of the pre-implementation, we inquired about the nurses' staff expectations regarding implementation, and one of the main issues was the time it would take to complete the nurse record, which motivated the present work. Based on the results of this study, the IMCU direct care and non-patient related tasks matched with the estimated percentage. However, there is an enormous difference in support and personal activities. In the GCU, there was difference in all the categories. In the ICU, the most significant differences were in indirect care and support.

The support task category had zero or few activities in all units. Because this is the first study on this subject, this category will be re-evaluated in a future study.

Except for in the IMCU (that had an almost equal proportion of the direct and indirect care tasks), there was a greater proportion of time spent in the indirect care activities than the direct ones, with less time spent in the personal activities.

Regarding daily activities, the IMCU was the only unit that showed the greatest variations over the course of the study. This may relate to the different patients' complexity of care, the nurse:patient ratio, and/or the infrastructure of the facility compared to the other units. On the other hand, at the end of the study in the ICU, there was a tendency to use the EHR more extensively, and to perform more direct care activities in the GCU. The work sampling method may reduce the Hawthorne effect, but in this first evaluation we cannot determine if this tendency is the product of this effect or a particular characteristic of both units.

Regarding EHR documentation, the percentage was greater than expected in all of the units. ICU showed the greatest percentage (23.79%). According to the literature, the nurse staff expends 15.79% (95% CI 14.25, 17.33) of their time in all documentation tasks, including paper documentation (10.55%) and EHR use (5.24%) [16]. Other authors mention 17.7% [9], 10.1% [17], and even 35.3% time commited to documentation. While there is certain agreement regarding the time modification when transitioning from paper to electronic documentation, it is unclear when there is an update or redesign to an existing electronic record system. For this reason, when the impact to the workflow during EHR implementation is recorded, some studies consider the time associated to registration and direct care as primary and secondary indicators [18].

We included "personal activities" as an individual category separate from the others (or included but in a different category like non-value-adding work [13]). The total time comprising this category contributes to the better understanding of how nurses use their time.

The findings of this study give us useful preliminary information regarding how nurses use time in completing assigned tasks. However, there are some limitations. The WS technique is one of the most frequently used due to its usefulness and cost effectiveness [19], and findings can be compared statistically. However, there are variations in the use of the technique [13]. Additionally, nursing task definitions do not always match in these studies. In spite of this, WS is a useful technique when there are constraints on time and resources, making the evaluation feasible.

We researched adult units during morning and afternoon shifts. To evaluate the night shift, it would be necessary to adapt the list of activities, due to the fact that there are different types of tasks performed during the night shift. Even though the observations took place during three weeks, the observed time only corresponds to five days. However, we obtained the required minimum of observations. To be able to determine if the implementation of the new EHR version will or will not affect the time, it is necessary to complete the evaluation in the post-implementation phase, planned six months after the implementation. It will be necessary to increase the number of units involved, and ideally should be done over a much longer period.

The data regarding the time spent by the nurses in their different tasks enables the evaluation of changes in the nursing practice after the implementation of a new system [9]. While this information is important, it only reflects one part of the phenomenon, since implementation may not only change the time used for doing a task but also the workflow and, consequently, the continuity of care. Further research will help to better understand this situation and to perform improvements during the process.

Acknowledgments

The authors would like to thank the students Leonel Cameselle, Agustina Bertoia, and Leonardo Juarez for tasks related to field observations, data entry, and translation of the manuscript. We extend special acknowledgements to the Chief of Nursing Department, Lic. Elsa Villalba, and the nursing personnel of HIBA.

This research was done under the Health Informatics Residence Program.

References

[1] Meadows G. The nursing shortage: can information technology help? Nurs Econ. 2002 Feb;20(1):46–8.

[2] Thompson D, Johnston P, Spurr C. The impact of electronic medical records on nursing efficiency. J Nurs Adm. 2009 Oct;39(10):444–51.

[3] Ward MM, Vartak S, Schwichtenberg T, Wakefield DS. Nurses' perceptions of how clinical information system implementation affects workflow and patient care. Comput Inform Nurs CIN. 2011 Sep;29(9):502–11.

[4] Ward MM, Vartak S, Loes JL, O'Brien J, Mills TR, Halbesleben JRB, et al. CAH staff perceptions of a clinical information system implementation. Am J Manag Care. 2012 May;18(5):244–52.

[5] Myny D, Van Goubergen D, Gobert M, Vanderwee K, Van Hecke A, Defloor T. Non-direct patient care factors influencing nursing workload: a review of the literature. J Adv Nurs. 2011 Oct;67(10):2109–29.

[6] Electronic Nursing Documentation: Charting New Territory [Internet]. [cited 2014 Dec 3]. Available from: http://www.medscape.com/viewarticle/810573

[7] Chase R. Operations management nella produzione e nei servizi. Milano: McGraw-Hill; 2012.

[8] Marriner-Tomey A. Guide to Nursing Management and Leadership. 6th Edition. St. Louis, MO: Mosby; 2000.

[9] Munyisia EN, Yu P, Hailey D. How nursing staff spend their time on activities in a nursing home: an observational study. J Adv Nurs. 2011 Sep;67(9):1908–17.

[10] Wolff J, McCrone P, Auber G, Fiedler P, Patel A, Wetterauer U. Where, when and what? A time study of surgeons' work in urology. PloS One. 2014 Jan;9(3):e92979.

[11] Davies JE, Barber N, Taylor D. What do community pharmacists do?: results from a work sampling study in London. Int J Pharm Pract. 2014 Oct;22(5):309–18.

[12] Wirth F, Azzopardi LM, Gauci M, Adami MZ, Serracino-Inglott A. Time and motion study for pharmacists' activities in a geriatric hospital. Int J Pharm Pract. 2009 Dec 8;17(6):373–6.

[13] Antinaho T, Kivinen T, Turunen H, Partanen P. Nurses' working time use - how value adding it is? J Nurs Manag. 2014 Oct 3;

[14] Sittig DF. Work-sampling: a statistical approach to evaluation of the effect of computers on work patterns in the healthcare industry. Proc Annu Symp Comput Appl Sic Med Care Symp Comput Appl Med Care. 1992 Jan;537–41.

[15] Gardner D, Gardner GA, Middleton S, Phillip D. The Nurse Practitioner Research Toolkit. 2009. p. 78.

[16] Korst LM, Eusebio-Angeja AC, Chamorro T, Aydin CE, Gregory KD. Nursing documentation time during implementation of an electronic medical record. J Nurs Adm. 2003 Jan;33(1):24–30.

[17] Cornell P, Herrin-Griffith D, Keim C, Petschonek S, Sanders AM, D'Mello S, et al. Transforming nursing workflow, part 1: the chaotic nature of nurse activities. J Nurs Adm. 2010 Sep;40(9):366–73.

[18] Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. J Am Med Inform Assoc JAMIA. 12(5):505–16.

[19] Pelletier D, Duffield C. Work sampling: valuable methodology to define nursing practice patterns. Nurs Health Sci. 2003 Mar;5(1):31–8.

Address for correspondence

Lic. Maria B. Schachner

bibiana.schachner@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-259

Web-tool to Support Medical Experts in Probabilistic Modelling Using Large Bayesian Networks With an Example of Hinosinusitis

Mario A. Cypko[a], David Hirsch[a], Lucas Koch[a,b], Matthaeus Stoehr[c], Gero Strauss[d], Kerstin Denecke[a]

[a] *Innovation Center Computer Assisted Surgery, University Hospital of Leipzig, Leipzig, Germany*
[b] *Usability Department, Merseburg University of Applied Sciences, Merseburg, Germany*
[c] *ENT Department, University Hospital Leipzig, Leipzig, Germany*
[d] *International Reference and Development Center for Surgical Technology, Leipzig, Germany*

Abstract

For many complex diseases, finding the best patient-specific treatment decision is difficult for physicians due to limited mental capacity. Clinical decision support systems based on Bayesian networks (BN) can provide a probabilistic graphical model integrating all necessary aspects relevant for decision making. Such models are often manually created by clinical experts. The modeling process consists of graphical modeling conducted by collecting of information entities, and probabilistic modeling achieved through defining the relations of information entities to their direct causes. Such expert-based probabilistic modelling with BNs is very time intensive and requires knowledge about the underlying modeling method. We introduce in this paper an intuitive web-based system for helping medical experts generate decision models based on BNs. Using the tool, no special knowledge about the underlying model or BN is necessary. We tested the tool with an example of modeling treatment decisions of Rhinosinusitis and studied its usability.

Keywords:
Clinical decision support system; Bayesian network; conditional probability tables; probabilistic modelling; expert system; treatment decision; web tool; rhinosinusitis.

Introduction

For some patient-specific treatment decisions (e.g., in oncology) more patient information needs to be considered at once than a single physician is able. For this reason, time intensive meetings of experts from different medical domains (e.g., surgery, radiology and radiotherapy) are conducted to understand and discuss the entire patient situation and possible therapy options and outcomes. However, each of the experts participating in these meetings considers the patient-specific situation from an individual viewpoint or background, making a discussion and finding a unanimous decision difficult.

Intelligent clinical decision support systems (CDSSs) [1] based on probabilistic graphical models, and more specifically, a Bayesian network (BN) [2], could support physicians by modeling complex interdisciplinary treatment decisions and simulating integrated decision making. A BN's graphical model contains nodes representing information entities (IE) with a set of events that can occur (e.g. Boolean values). In the domain of medicine, IEs describe for example results regarding medical examinations, medical imaging, patient's compliance, genetic factors, or describe patient characteristics (e.g. age, gender, tobacco and alcohol consumption). Nodes are linked by directed edges. More specifically, a parental node is connected to a child node by edges representing their direct causality.

The probabilistic model is dependent on the modeled graph structure and represents the strength of the causality between an IE and its direct linked causes (parental nodes) by a conditional probability table (CPT). In a CPT, for each event of an IE probabilities are assigned based on all permutations of its parental events. Consequently, the amount of necessary probabilities for a CPT grows exponentially with the number of parental nodes and their events. A main challenge when modeling the graphical structure of BNs is to find the right balance between the granularity of IEs with their events and the complexity of the model in order to avoid large CPTs. Based on our previous work [3], CDSS for increasingly complex treatment decisions requires more detailed BN models.

The graphical model of a BN can be created by applying machine learning algorithms to a set of data collected from guidelines and other sources that define relations between IEs. The information already provide the probabilities [4-6]. Our previous study revealed a significant disadvantage of this method for complex diseases, evidences for many IEs are usually based on easier accessible and cheaper patient information (e.g. age, gender, tobacco- and alcohol consumption) [3]. Thus, proven relations as they are available in guidelines and statistics do not represent the desired natural direct dependencies in the graphical model.

In contrast to machine learning, our approach creates the model structure by considering the natural direct dependencies without limits given by statistics. The conditional probabilities are assigned in a subsequent step. In that way, the graphical model forms the basis for collecting probabilities. This kind of model-based medical evidence [7] is expensive to achieve at the moment, and manual modeling by medical experts is subjective and very time-consuming.

To overcome this limitation, we developed a web-based tool that allows experts to assign probabilities to nodes and respective events in a graphical model describing treatment decisions. In this paper, we introduce the system. It was tested on a the graphical BN model representing treatment decisions related to acute- and chronical rhinosinusits (ARS and CRS). The paper is structured as follows: In the *Methods* section, we describe the web tool to collect the CPTs for the probabilistic model. We evaluated the tool by having physicians assign probabilities to events in a treatment decision model. In the section *Evaluation and Results,* we present a comparative

assessment of the expert's probability values by using an intraclass correlation, and by analysing metadata recorded during the assignment process. Further, we present the results from a usability study. In the *Discussion* section, we describe advantages and disadvantages of the application, reasonable issues for the disparities in the expert's assessments and necessary extensions. The *Conclusion* summarizes the results and provides an outlook on future expert-based modeling in the medical domain.

Methods

Web-based Tool for Assigning Probabilities

Previous work from L.C. van der Gaag et al. [8] showed that experts can assign probabilities to IE in a given probabilistic network by formulating natural language questions and allowing the assessments to be provided on a scale. When prompted, domain experts were able to provide probabilities at a rate of over 150 probabilities per hour. Our approach is based upon that work, but goes beyond by using a computer system to digitalize the process of eliciting probabilities from domain experts through a web-based tool.

Our CPT-tool runs as a server application on the Node.js architecture and uses MongoDB as a persistence layer. The application requires a user authentication to prevent misuse of the system by unauthorized parties, and also to record user specific metadata that is used for the evaluation of the system. As a web application, the tool can be used from any browser and is thus system independent.

The CPT-tool takes as input a probabilistic network in XML format that contains all nodes of the model with its events and edges from parental nodes. We are building these networks using the open source software UnBBayes [9], but any other tool for generating BN models would be applicable. The CPT-tool automatically extracts the nodes, their events, and edges from the model to *(1)* generate an overview table and two types of questionnaires, *(2)* generate CPT specific questionnaires of permutations of parent events and *(3)* node specific selection page of probability combinations. The output of the application is again an XML file stored in the same XML format as the input, but with the probabilities set by the user.

User Interactions with the System

In the following, we describe the interactions between medical experts and the system when assigning probabilities to a graphical BN model. After logging into an account, a user can upload a probability network with events to which probabilities need to be assigned.

(1) Once the network has been analyzed by the system, the pertaining nodes are displayed in an overview table to the user, see figure 1. Each row contains the number of pages the domain expert has to complete (i.e. the number of permutations of the given node) and the number of variables each node contains. The domain expert is able to freely choose which node to start the elicitation process, by clicking the corresponding button next to the node. The required probabilities can be set at any time and from anywhere, which doesn't force the domain expert to a fixed schedule that might otherwise cause an inconvenience.

(2) The first questionnaire is similar to the elicitation sheet suggested by van der Gaag [8]: The selected IE is presented as a text fragment, but is divided into two parts. One part presents the preconditions of the current iteration, while the other part presents the events of the current node to be assessed. For example, figure 2 asks the user to assess the probability of antibiotics being used, given the state of the parent events on the left hand side. Since the events of the antibiotics node are either true or false, the user is only required to set the probability for the question: *How likely is it that antibiotics is: true.* The tool specifies the percentage bar in seven percent steps (1%, 15%, 25%, 50%, 75%, 85%, and 99%) with a text description ("(almost) impossible", "improbable", "uncertain", "fifty-fifty", "expected", "probable", and "(almost) certain") assigned to each of the percentages. However, there is still the possibility to set the exact percentage in 1% increments.

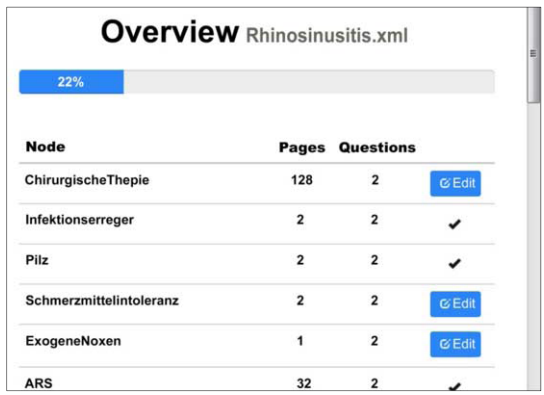


Figure 1–Overview table from CPT-tool

In case of determining the probabilities of a node with more than two events (e.g. Tumour-state), the expert is asked to set the probabilities with a total of 100%. At the end of the page, a multiple choice selection to evaluate the confidence level of the assessments is presented. The options are: very confident, confident, unconfident and very unconfident. If the assessment was unconfidently made, an additional input field is presented to describe the uncertainty of the evaluation. After submitting the probabilities, the next permutation of the node's parent events will be presented. But before, the system interpolates all probabilities to 100% in the background based on the combination of parent events, by dividing each of these probabilities by their quotient. Their quotient is the sum of the probabilities, divided by 100.

For example: The probabilities for the events $\{n_1, n_2, n_3\}$ of the node n are set with 35%, 30%, 60%. Then, the system divides each of the probability by 125%, because of (35 + 30 + 60) / 100 = 125%. 28% + 24% + 48% = 100%.

At the top of the page, the user is presented with a progress bar guiding her through the node iterations. After assessing all probabilities of a node, the user is returned to the node overview page where the recent node will be marked as completed.

(3) If a chosen node n has at least two parent nodes, before step (2), an additional questionnaire is presented to find significant combinations of parental events, see figure 3. On this page, the probability of n dependents on a non-empty and incomplete set of parent events E are to be asseessed, so that the probability of n is set for all CPT items that contain the combination of E. Significant combination of parental events are identified using the following hypothesis:

The probabilities for a node *n* are the same, if an event or a combination of events *E* of *n*'s parental nodes occurred independent on all combinations of events of all other unconsidered parental nodes of *n*. ■

The percentage for the node dependent on the selected combination is assessed by the user in the same way as in the other questionnaire (step 1). After confirming a combination with a percentage, the combination is visualized by a blue dot with a number. Same numbers indicate one combination and the number itself the order of setting the combinations.

Figure 2– Example of the questionnaire from CPT-tool

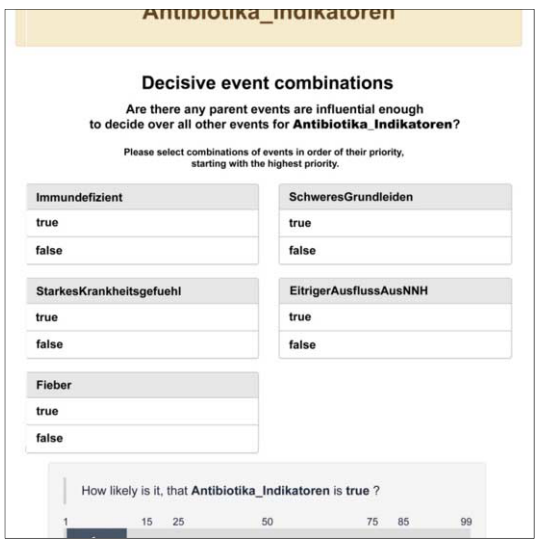


Figure 3– Questionnaire for significant combinations of parent events

Finally, after completing the assessment of all nodes, the user can export the network encompassing the elicited probabilities for each node.

Evaluation Methodology For evaluating the CPT-tool, we performed a user study with clinical experts, which also included a usibility study. Evaluation methods and results are described in the following.

Study Design: CPT-tool Used by Medical Experts

For a study of the presented CPT tool, we used a graphical model for the treatment decision of acute- and chronic rhinosinusits (ARS, CRS) comprising 75 nodes and 100 dependencies. The model was handcrafted, first by non-medical-experts based only on guidelines. For graphical representation and later inference tests, the model was implemented using UnBBayes. Three physicians were asked to correct and validate the graphical model and later, to set the CPTs by using the presented web tool. One physician was a resident physician in his second year and experienced in Bayesian theory. The other two were ENT-surgeons with 7 to 10 years of experience of ARS and CRS, with around 1000-1200 treatments a year, from this around 520 to 780 of which are surgical treatments and the remaining are conservative therapies.

The CPT-tool computed for the rhinosinusitis model contained 1526 questions, i.e., the number of probabilities that needed to be assigned. For assigning the probabilities, accounts for all three physicians were created. The experts were free to choose when and where they would use the CPT-tool to answer the questions. The only requirements were to finish a node once it was started, to keep a general overview of the node in mind, and to finish the whole assignment in two weeks. The second questionnaire for significant combinations was not part of this first study, but was assessed at a later time by resident physician.

During the assignment process, meta-data was collected to help with the evaluation of the users' decision making process. These include the previously mentioned self-evaluation of the users' choices, the amount of time it took for a user to set a probability and complete a node, and the order in which the nodes were completed. The purpose of this evaluation was to study the time required for probability assignment, to study the user satisfaction with the CPT-tool and to find out to what extent the experts agree in probability setting.

Results

Usability Study: Design and Objectives

After the first study, and based on the experiences, a usability study with 20 participants of different gender and ages between 25 and 44 years was done. Focus of the usability study was to identify potential sources of error that may lead to wrong assessments caused by the application's construction, design and usability. Therefore, a special, smaller model was created with a more intuitive example about a common topic, the accident rate of traffic, with IEs such as *road surface, the type of road, weather and season conditions*. For the tests, different usability methods were combined to get qualitative and quantitative results out of the study, such as screen capturing, voice recording, eye tracking, questionnaires and interviews. During this study, the participants were asked to think out loud and to answer questions about their actions. The participants were asked to use the CPT-tool independently by their own intuition but with the possibility to ask the usability experts if necessary.

Evaluation Results

The assignment process of the rhinosinusitis probability network took each expert an average of 7 to 12 seconds per elicitation step. At just over 1500 probabilities to assess, the experts were able to complete their task at a rate of 300 to 500 probabilities per hour, with a minimum of 2 seconds for the fastest and 26 seconds for the slowest answer. This means that a total time of 3.5 to 5 hours was necessary to complete the elicitation process for this particular model. With the aid of selecting decisive event combinations, the resident physician was able to eliminate 47.97% (n = 732) of all probabilities using less than 50 assessments.

From the metadata, we could recognize that the experts started with the nodes that had the least number of questions, before working through the remaining nodes in chronological order. Comparing the results of our experts, we found that only 148 out of the 1500 probabilities had a deviation of 50% or more. However, 501 probabilities had a deviation of over 15%. To better evaluate the assessments of the experts, we employed intraclass correlation as a form of reliability testing. The intraclass correlation coefficient kappa describes how closely the elicited values resembled each other in six categories: no agreement, slight, fair, moderate, substantial, and almost perfect agreement [10]. Figure 4 shows the kappa, on a scale of -1 to 1, for each node that has at least one parental node. Nodes without parents are ignored based on their opinionated characteristic. We did not specify in the assignment rules whether probabilities to such nodes have to be set on the basis of the specific epidemiology. The graph depicts an overall positive result, with an average kappa of 0.16, indicating "moderate agreement". Twelve of the thirty nodes depicted in the graph were categorized "substantial agreement". The few nodes found during the intraclass correlation reliability test that fall into the "fair agreement" category will be discussed and reevaluated in an upcoming post-elicitation meeting with the experts.

The two experienced experts in our study did not make use of the applications feature to report uncertainties during the elicitation process. Instead, only the resident provided feedback on several of his decisions, most frequently criticizing the epidemiology of the nodes. At this stage, it was not clear if these deviations in probabilities and the almost positive assessments of confidence originate from the differing medical expertise or is caused by the usability of the system. For this reason, we performed a usability study.

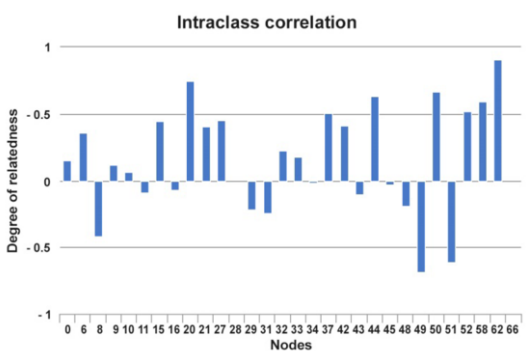


Figure 4– kappa for each node with at least one parental node from the rhinosinusitis model

In the usability study, we recognized 28 negative aspects determined by the system that we classified in three levels of severity: (1) critical – system crash or misleading to wrong probabilities (6 of 28). (2) user-unfriendly – demotivating or loss of attention (6 of 28), and (3) nice-to-have – is not affecting the rating (16 of 28). Table 1 shows the problems and the number of affected testers from the severity level 1 and 2. The additional step of presetting key combinations that was integrated only in the usability study was one of the main misleading problems. Problems with the severity level 2 lead to a longer answering time. For the most participants the study has shown first signals of demotivation after 15 to 30 minutes; after around 30 to 45 minutes participants started to loose attention. The third level is not listed as it contains subjective feelings. Only a few testers were affected by the same aspects and they did not influence the usability of the tool.

Table 1– Table of negative aspects determined by the system

Problems	# affected testers out of 20
Level of severity 1:	
Understanding of pre-setting key combinations of causes unclear	12
Group of direct causes not recognized	7
Duration of the questionnaire for one node too long	7
Only the first probability bar taken serious	5
Relation between all bars unclear	4
How to interpret the input field for uncertainty unclear	4
Some web browsers allow to stop pop-ups which prohibits functions of the tool	3
Level of severity 2:	
Comment feature hinders the assessment process	4
The node terms or subject unclear – missing instructions	3
Duration of "pages" and "questions" unclear	1
Information on lost data when closing the application	1
Program entraps to clicking through the questionnaire	1
Not enough application feedback, "unable to tell what's happening in the background"	1

Discussion

In the following , the results of both studies are discussed with examples for a better usability of the CPT-tool, and also compared with previous literature. To finish the first study with the rhinosinusits model, a follow up meeting with the experts will be needed for personal interviews and to evaluate and discuss any problems or difficulties that occurred during the elicitation process. Additionally, nodes that have led to differing assessments will be discussed to compile a convergent model for further evaluations.

The usability study with 20 participants gave highly reliable result. Early researches of Nielson and Landauer [11] showed, that five users will find about 85% of all existing usability problems and "when collecting usability metrics, testing 20 users typically offers a reasonably tight confidence interval" [12]. Based on the results from both studies, all problems with a severity level 1 should be corrected and also some of level 2 with higher numbers of affected testers. Especially, the additional step of presetting key combinations is a powerful and necessary approach for treatment decision models to minimize the large amount of probabilities. By addressing the issues revealed during the usability study, the CPT-tool can be improved to build an intuitive tool that allows domain experts to help in collecting probabilities without any need of knowledge about the underlying network or BNs. This could lead to a collection of large amount of CPTs in a short amount of time by using the crowdsourcing principle, so that later a justifiable average of probabilities can be composed.

Another significant problem is that the duration of setting the CPT for a node without interruptions is too long and leads to impatient, inattentive, fast and incorrect assessment. This problem can be solved in many ways, for example by using the average time of expert based answering, multiplied with the number of a node's questions the expert will have a better feeling for the upcoming effort. Also, the interruption during the probability setting of a node could be allowed by saving the current editing state, but with the resumption of the node its presetting history should be accessible and studied by the expert to guarantee an overview on the graph. Eliminating the problems, we expect a very intuitive tool. Another study of the upgraded application will follow.

Our CPT-tool has several advantages compared to the method from van der Gaag [8], in particular with respect to flexibility, answering rate, and precision. The separation of current and dependent events allows the user to see the preconditions all in one place, and also enables the questionnaire and underlying network to take on any form and size. With a time consumption of approximately 11 seconds per answer, the average answer rate is substantially higher. A digital solution also allows to present an exact percentage. Although it is known from van der Gaag's work that the precision of 1% increments is not necessary for the CPTs, it was one of our expert's requests to be able to distinguish between similar answers. An important and promising solution for large models is the questionnaire of significant combinations. At this stage, the usability study shows that this questionnaire is not intuitive or obvious, but it can significantly decrease the amount of questions. In a BN example with laryngeal cancer [13] some nodes in the model contain a CPT with more than ten thousands of probabilities, but can be minimized to only a small number of 20 to 30 nodes by presetting key combinations.

At the moment, the CPT-tool uses as input an XML-format given by UnBBayes's export-file that is very extensive and software specific. In the next version of the CPT-tool, this will be changed to a simpler XML-format, so that also other software export could be parsed and used in the web-tool. Additionally, an upgrade for instantiable models as multi-entity Bayesian network is planned.

Conclusion

In this paper, we introduced a web-based tool for assigning probabilities to BN models of clinical treatment decisions. The tool was evaluated with respect to usability and in a user study. The assessments we have received from the domain experts are promising even though we collected problems of three level of severity. At least level 1 problems need to be eliminated to minimize the number of wrong expert based assessments caused by the usability of the presented web-tool. The usability study not only increased the understanding of usability issues but also improved the development of an intuitive web-tool. This is a major step in encouraging medical experts to use probabilistic modeling. Based on this work, another web-tool for creating graphical models based on natural language questions is planned. These kinds of tools will allow crowdsourcing the development of CDSSs.

Acknowledgments

This research has been done within the research group "Digital Patient and Process Modeling" funded by the German Ministry of Research and Education.

We would like to thank L. Koch's usability group of the University of Merseburg for a professional and volunteered evaluation of our tool. We also like to thank Prof. Strauss's AQUA clinic team of medical experts and the International Reference and Development Center for Surgical Technology (IRDC), and also the ENT department of the University Hospital of Leipzig.

References

[1] Berner ES. Clinical Decision Support Systems: State of the Art. Agency of Healthcare Research and Quality, 2009.

[2] Pearl J. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, 1988.

[3] Stoehr M, Cypko M, Denecke K, Lemke HU, and Dietz A. A model of the decision-making process: therapy of laryngeal cancer. Int J CARS; 9, 1: 217-218, 2014.

[4] Cooper GF, and Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning; 9, 4: 309-347, 1992.

[5] Friedman N, Geiger D, and Goldszmidt M. Bayesian Network Classifiers. Machine Learning; 29, 2-3: 131–163, 1997.

[6] Getoor L, and Taskar B. Introduction to Statistical Relational Learning. MIT, 2007.

[7] Lemke HU, and Berliner L. Personalised medicine and patient-specific modelling. In Personalisierte Medizin, 14. Edition, Dresden: Health Academy; 55–164, 2010.

[8] Van der Gaag LC, Renooij S, et al. Probabilities for a probabilistic network: a case study in oesophageal cancer. Artif Intell Med 2002; 25: 123-48, 2002.

[9] Carvalho RN, Santos LL, Matsumoto S, Ladeira M, and Costa PCG. UnBBayes-MEBN: Comments on Implementing a Probabilistic Ontology Tool. In IADIS Applied Computing 2008 conference, 2008.

[10] Landis JR and Koch GG. The measurement of observer agreement for categorical data. Biometrics, 33:159174, 1977.

[11] Nielsen J, and Landauer TK. A mathematical model of the finding of usability problems, Proceedings of ACM INTERCHI'93 Conference, pp. 206-213, 1993.

[12] Nielsen J. Quantitative Studies: How Many Users to Test?. Nielsen Norman Group, 2006. Retrieved 14 December 2014 from http://www.nngroup.com/articles/quantitative-studies-how-many-users/.

[13] Cypko M, Stoehr M, Denecke K, Dietz A, and Lemke HU. User interaction with MEBNs for large patient-specific treatment decision models with an example for laryngeal cancer. Int J CARS; 9, 1, 2014.

Address for correspondence

Mario A. Cypko

Innovation Center Computer Assisted Surgery

University of Leipzig, Semmelweisstr.14, 04103 Leipzig, Germany

mario.cypko@medizin.uni-leipzig.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-264

Physicians' Attitudes Towards the Advice of a Guideline-Based Decision Support System: A Case Study With OncoDoc2 in the Management of Breast Cancer Patients

Jacques Bouaud[a,b], Jean-Philippe Spano[c,d], Jean-Pierre Lefranc[c,e], Isabelle Cojean-Zelek[f], Brigitte Blaszka-Jaulerry[g], Laurent Zelek[h], Axel Durieux[i], Christophe Tournigand[j], Alexandra Rousseau[k], Pierre-Yves Vandenbussche[l], Brigitte Séroussi[b,m]

[a] AP-HP, DRCD, Paris, France
[b] Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, LIMICS, Paris, France ; INSERM, U1142, LIMICS, Paris, France ; Université Paris 13, Sorbonne Paris Cité, LIMICS (UMR_S 1142), Bobigny, France
[c] Sorbonne Universités, UPMC Université Paris 06, UFR de Médecine, Paris, France
[d] AP-HP, Hôpital Pitié-Salpétrière, Service d'Oncologie médicale, Paris, France
[e] AP-HP, Hôpital Pitié-Salpétrière, Service de Chirurgie et Cancérologie gynécologique et mammaire, Paris, France
[f] Hôpital des Diaconesses, Pôle oncologie médicale, Paris, France
[g] CH Lagny Marne la Vallée, Service de Radiothérapie-Oncologie, Lagny, France
[h] Université Paris 13, UFR SMBH, Bobigny, France; AP-HP, Hôpital Avicenne, Service d'oncologie médicale, Bobigny, France
[i] Institut de Cancérologie des Peupliers, Paris, France
[j] AP-HP, Hôpital St-Antoine, Service d'oncologie médicale, Paris, France
[k] AP-HP, Hôpital St-Antoine, URC-EST, Paris, France
[l] Fujitsu (Ireland) Limited, Swords, Co. Dublin, Ireland
[m] AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France; APREC, Paris, France

Abstract

When wrongly used, guideline-based clinical decision support systems (CDSSs) may generate inappropriate propositions that do not match the recommendations provided by clinical practice guidelines (CPGs). The user may decide to comply with or react to the CDSS, and her decision may finally comply or not with CPGs. OncoDoc2 is a guideline-based CDSS for breast cancer management. We collected 394 decisions made by multidisciplinary meeting physicians in three hospitals where the CDSS was evaluated. We observed a global CPG compliance of 86.8% and a global CDSS compliance of 75.4%. Non-CPG compliance was observed in case of a negative reactance to the CDSS, when users did not follow a correct CDSS proposition (8.6% of decisions). Because of errors in patient data entry, OncoDoc2 delivered non-recommended propositions in 21.3% of decisions, leading to compliances with CDSS and CPGs of respectively 21.4% and 65.5%, whereas both compliances exceeded 90% when CDSS advices included CPG recommendations. Automation bias, when users followed an incorrect CDSS proposition explained the remaining non-compliance with CPGs (4.6% of decisions). Securing the use of CDSSs is of major importance to warranty patient safety and benefit of their potential to improve care.

Keywords:

Clinical decision support systems, Attitude to Computers, Guideline adherence, Breast cancer management.

Introduction

Variations in medical practices have been observed for decades around the world. If some level of variation related to differences in population need, health status or patient preferences, is expected, "unwarranted variation" raises questions about quality and appropriateness of care [1]. Following evidence-based medicine principles, clinical practice guidelines (CPGs) have been developed to eliminate these "unwarranted variations". Recommended by health professional societies or national health agencies, CPGs are information resources describing the recommended therapeutic management for different clinical situations encountered in given pathologies. They are thus intended to improve the quality of clinical care, reduce inappropriate variations, produce optimal patient outcomes, and promote cost-effective practices.

Many studies have shown that the sole dissemination of narrative CPGs had nearly no impact on physician behavior. On the opposite, numerous reviews [2] suggest that clinical decision support systems (CDSSs) have the potential to promote CPGs use. However, many studies showed positive effects of CDSSs, whereas others found only a limited impact of these systems upon physicians' behavior. One of the difficulty comes from the wide, non-totally agreed upon, meaning of CDSSs. This denomination covers indeed systems ranging from the simple provision of brief prompts and calculation services, through the provision of the therapeutic support for the management of complex patients with multimorbidity. CDSSs may also vary according to user interaction modalities, operating either as automated reminder systems generating alerts, or as on-demand systems. A recent study [3] reported for instance that presenting decision support within electronic charting or order entry system was associated with failure compared with other ways of delivering advice. In addition, requiring practitioners to provide reasons when overriding propositions and providing advice concurrently to patients and practitioners were more likely to be effective.

Cancer management is subject to practice variations and to varied levels of compliance with oncology CPGs [4].

Multidisciplinary meetings (MDMs), or tumour boards, are generally associated with improvements of guideline compliance rates [5]. However, the efficiency of MDMs has been recently questioned [6] and CDSSs used in MDMs are considered to improve the compliance of MDM decisions with CPGs [7].

OncoDoc2 is a guideline-based decision support system applied to the management of non-metastatic breast cancer patients to be used by MDM physicians [8]. Routinely used for three years in breast cancer MDMs of the Tenon hospital (Paris, France), it showed very good results in terms of compliance [9]. More recently, we conducted a prospective cluster randomized controlled trial to evidence the impact of the system on the compliance of MDM decisions with CPGs. When studying the sole intervention arm of the trial, where MDM physicians used the system for all breast cancer patients, we observed that, in half of the cases, the system was not correctly used. In theses cases, patients were described with "errors", and the compliance rate of MDM decisions was significantly decreased. We concluded that it was better not to use the system than to use it improperly [10].

Indeed, health information technology is expected to improve the quality of care. Tools such as electronic healthcare records (EHRs), CDSSs, computerized provider order entry systems (CPOEs), are believed to make the delivery of healthcare safer, more effective and more efficient. However, with the widespread implementation of HIT, studies that suggest HIT introduces unpredicted and unintended adverse consequences (UACs) potentially harmful for patients have been published. Campbell et al. [11] have identified 9 types of UACs, among which the Type 7 category (new kinds of errors) covers the concept of "e-iatrogenesis" defined by Weiner et al. [12] as "patient harm caused at least in part by the application of health information technology". These errors, known as technology-induced errors, are the result of a mismatch between the functioning of HIT tools and the real-life demands of healthcare work. Sittig and Singh [13] reported that UACs of HIT occur when HIT is unavailable for use, malfunctioning during use, or used the wrong way. Usability engineering methods are thus increasingly being used to ensure improved system usability and the safety of HIT applications [14].

When studying the non-compliance of physician decisions with *state-of-the-art CPGs* while using a guideline-based CDSS, e-iatrogenesis is not always involved. Physicians may not comply with the system propositions and not decide according to the best evidence for the patient. But they may also *comply* with the system propositions and not decide according to the best evidence for the patient. Two main scenarios should be considered. In the first one, the CDSS is correctly used and propositions are "correct" but physicians do not follow them. This corresponds to the mechanism of "reactance" borrowed from cognitive engineering, and introduced by Vashitz et al. [15]. Psychological reactance is an unpleasant motivational state, in which people react to situations, they feel their autonomy is threatened. This is the negative side of reactance, in which physicians want to reaffirm their freedom of choice and consciously or unconsciously either ignore the CDSS propositions or choose on purpose a different course of action. In this case, the quality of the resulting decision has deteriorated as compared to the system propositions. This negative side is balanced by the positive reactance, in which physicians do not follow the system propositions but built on them to refine and thus improve their decision according to the specific condition of the patient or to take into account the very last published evidence not yet integrated in the CDSS knowledge base. In the second scenario, we are in the case of the automation bias (AB) [16], i.e. the tendency to over-trust HIT leading a physician to make an incorrect decision in order to follow the advice provided by a CDSS [17]. Automation bias may be demonstrated by "negative consultation", a term used to denote when a correct decision is changed to an incorrect one on the basis of an incorrect advice. In a systematic review [18], negative consultations range from 6 to 11%.

The aim of this work is to analyze the intervention arm of the randomized controlled trial where OncoDoc2 had to be used for all breast cancer patients discussed during MDMs. In previous works with the same context, we studied the accuracy of actual data entry, and how the way OncoDoc2 was used had an impact on the quality of MDM decisions assessed by their compliance with CPGs [10]. In this paper, we study the attitude of MDM physicians towards the advice of OncoDoc2 in order to assess the part of reactance and automation bias when they finally make their decisions. This paper does neither discuss the methods nor present the results of the trial, which will be presented in a different paper.

Materials and Methods

The guideline-based CDSS OncoDoc2

OncoDoc2 [8] is a computerized guideline-based CDSS. It provides patient-specific recommendations for breast cancer patients according to CancerEst (local) CPGs. The system can be automatically run on patient data extracted from EHRs. It can also be used according to the document-based paradigm of decision-making where the knowledge base is interactively browsed by the user. In this case, using OncoDoc2 consists of answering, by a simple click on the right value, a sequence of closed-ended questions displayed in the interface to finally access the guideline-based advice of the system made of a set of recommended care plans. The user-controlled navigation through the knowledge base, noted N, instantiates criteria according to patient-specific data, and selects the theoretical clinical profile of the knowledge base that best fits the actual patient. Figure 1 shows a screenshot of OncoDoc2 displaying a navigation, characterized by the "Recapitulative", and the corresponding advices ("CancerEst Recommended Treatment Plans"). OncoDoc2 has been routinely used at the Tenon hospital (Paris, France) with a compliance rate of 91.7% [9].

Data collection

We have analyzed the data produced by the weekly organized MDMs of the three hospitals of the intervention arm of the randomized controlled trial. In each of the three hospitals, MDM physicians had to use OncoDoc2 for all breast cancer patients. For each decision, we automatically collected the navigation performed by MDM physicians N^{MDM}, the advices provided by OncoDoc2 $\{R\}^{CDSS}$, and the MDM decision d. Additionally each week, clinical research assistants (CRAs) collected the list of patient cases discussed in MDMs the previous week. For each of them, they ran OncoDoc2 and performed their own navigation on the basis of the information found in patient medical records, without knowing what were the navigations of MDM physicians for the same patients.

The two types of compliance

We considered that for each patient, the CRA-performed navigation was the "gold standard" named "reference navigation", denoted N^{Ref}, and that the advices of OncoDoc2 attached to the reference navigation represented the set of reference guideline-based recommendations that applied to the

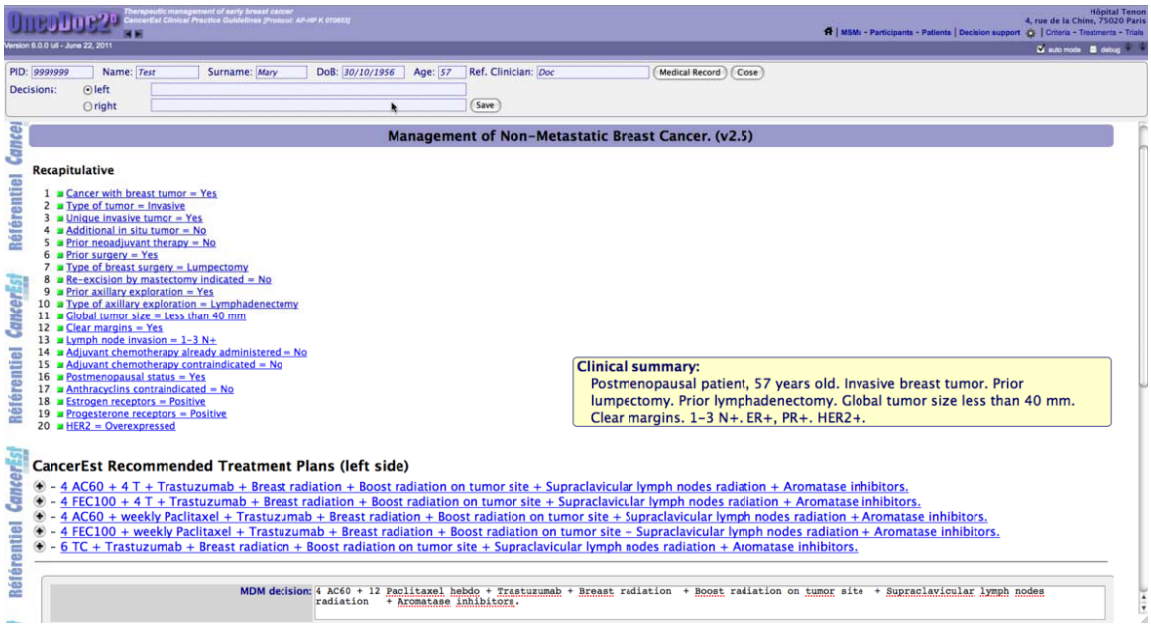


Figure 1 – A screenshot of OncoDoc2 providing advices as treatment plans for a given patient profile

patient, noted $\{R\}^{Ref}$. We stated that a MDM decision d complies with CPGs when $d \in \{R\}^{Ref}$, noted CPGs+.

When MDM physicians used OncoDoc2, and obtained the system advices $\{R\}^{CDSS}$, they could either follow one of the CDSS advices, or react and decide a different care plan. We state that a MDM decision d complies with OncoDoc2 when $d \in \{R\}^{CDSS}$, noted CDSS+.

Quality of the CDSS advice

As previously mentioned, misuses or errors when using CDSSs could lead to the delivery of non-CPG-compliant advices by the CDSS. In the context of OncoDoc2, errors in data input during the user-controlled navigation lead to situations in which $N^{MDM} \neq N^{Ref}$. Usually, in these situations, the set of advices provided by OncoDoc2, $\{R\}^{CDSS}$, does not correspond to the set of reference guideline-based recommendations $\{R\}^{Ref}$. However, it may happen that erroneous data entries lead to appropriate reference recommendations. When considering MDM decisions made with OncoDoc2, we distinguish three different situations:

- Advice+: $\{R\}^{CDSS} = \{R\}^{Ref}$. The CDSS proposes exactly the set of recommended care plans. This is systematically the case when the navigation N^{MDM} is correct, i.e. $N^{MDM} = N^{Ref}$, but can also be true when N^{MDM} is erroneous.
- Advice+/-: $\{R\}^{CDSS} \cap \{R\}^{Ref} \neq \emptyset$. The CDSS proposes at least one of the recommended care plans. N^{MDM} navigations are also necessarily erroneous.
- Advice-: $\{R\}^{CDSS} \cap \{R\}^{Ref} = \emptyset$. The CDSS proposes none of the recommended care plans. N^{MDM} navigations are necessarily erroneous.

Physicians' attitudes towards the CDSS advice

Considering the two types of compliance, CPGs+ and CDSS+, four different CDSS-supported decision-making attitudes, have to be considered. Notations are summarized in Table 1:

Table 1 – Summary of notations and categories of physicians' attitudes in CDSS-supported decision-making

CPGs+	Decision complies with CPGs
CPGs-	Decision does not comply with CPGs
CDSS+	Decision complies with CDSS
CDSS-	Decision does not comply with CDSS
Advice+	All CDSS propositions comply with CPGs
Advice+/-	At least one CDSS proposition complies with CPGs
Advice-	None of CDSS propositions comply with CPGs
Comp+	User follows CDSS and decision is CPG-compliant
React+	User does not follow CDSS and decision is CPG-compliant
Comp-	User follows CDSS and decision is not CPG-compliant
React-	User does not follow CDSS and decision is not CPG-compliant

- Comp+: CDSS+ $\wedge$ CPGs+. The user follows the CDSS advices and her decision adheres to CPGs, which corresponds to a positive compliance with the CDSS.
- React+: CDSS- $\wedge$ CPGs+. The user does not follow the advices of the CDSS, but her decision complies with CPGs. She reacts to the system to improve the final decision which corresponds to a positive reactance.
- Comp-: CDSS+ $\wedge$ CPGs-. The user follows the advices of the CDSS although they do not comply with CPGs. She does not have a critic attitude towards the system propositions and decides as proposed by the CDSS, which corresponds to a

negative compliance with the CDSS, also named automation bias.

- React-: CDSS- ∧ CPGs-. The user does not follow the advices of the CDSS and she does not decide according to CPGs. She reacts to the CDSS propositions to adopt a decision which does not comply with CPGs, which corresponds to a negative reactance.

Consolidation of the CDSS use and physicians' attitudes

We analysed the relationships between the type of advice delivered by OncoDoc2 and the type of physicians' attitudes depending on their compliance with CPGs and compliance with the CDSS. Figure 2 illustrates the possible configurations.

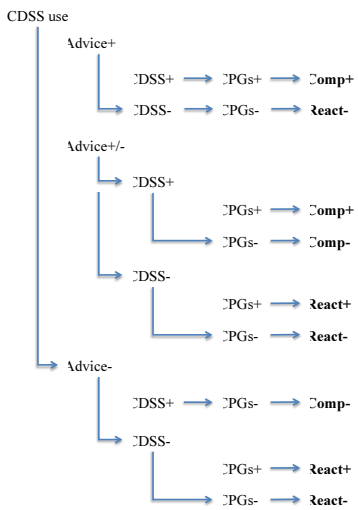


Figure 2 – Formalization of physicians' attitudes towards the CDSS depending on the type of the advice

Results

Data was collected for the intervention arm of the trial, including three hospitals, between June 2009 and April 2010. A total of 394 MDM decisions were made with the support of OncoDoc2 and considered for the analysis. The overall CDSS compliance rate was measured at 75.4%, showing a reactance rate of 24.6%. The overall CPG compliance rate was 86.8%.

As reported in [10], OncoDoc2 was incorrectly used in 52% of the cases. However, on the basis of the comparison with the recommendations of reference navigations, we observed that 68.8% of MDM decisions were made in Advice+ situations where all OncoDoc2's propositions were correct recommendations, and that 21.3% of MDM decisions were made in Advice- situations where all advices provided by OncoDoc2 were incorrect. Compliance with CPGs was significantly different according to the advice category (χ^2, $p < 10^{-9}$) and was measured at 65.5% for Advice-, whereas it was 92.3% for Advice+, and 94.9% for Advice+/-. In the same way, the CDSS compliance rate was significantly different according to the advice category (χ^2, $p < 10^{-37}$), and measured at 92.3%, 74.4%, and 21.4% for Advice+, Advice+/-, and Advice- respectively. Table 2 summarizes the distribution for the three advice categories and the associated CDSS and CPG compliances.

Table 2 – Distribution of advice categories and their corresponding CDSS and CPG compliances

Categories	n	(%)	CDSS+	CPGs+
Advice+	271	(68.8%)	92.3%	92.3%
Advice+/-	39	(9.9%)	74.4%	94.9%
Advice-	84	(21.3%)	21.4%	65.5%
Total	394	(100.0%)	75.4%	86.8%

The distribution of MDM physicians' attitudes towards OncoDoc2 is reported in Table 3. In 70.8% of the cases, the compliance with the CDSS advice was positive (Comp+), yielding CPG-compliant decisions. Reactance to the system advice represents nearly one quarter of all decisions (24.6%). Reactance is positive in about 2/3 of reacted decisions, leading to CPG-compliant decisions. Finally, the automation bias, or negative compliance with the system, is estimated at 4.6% of all MDM physician decisions. It must be noted that automation bias was observed only in the Advice- category.

Table 3 – Distribution of the four categories of MDM physicians' attitudes towards OncoDoc2 for decision-making

User attitudes	n	(%)
Comp+	279	(70.8%)
React+	63	(16.0%)
React-	34	(8.6%)
Comp-	18	(4,6%)
Total	394	(100%)

Discussion

OncoDoc2 was associated with a global CPG compliance of 86.8% and with a CDSS compliance of 75.4%. The expected effect of the CDSS, i.e. the positive compliance Comp+, when the 2 compliances simultaneously occur, was 70.8%.

More specifically, when CDSS advices include guideline-based recommendations (Advice+ and Advice+/-), CPG-compliance exceeds 92%. This suggests that the correct reminder of CPGs by a CDSS may foster their adoption. We observed that Advice- situations occurred in 21.3%. Such situations necessarily occur when the patient characterization in the system (N^{MDM} navigation) is not accurate. This proportion of situations where no guideline-based recommendation is provided by OncoDoc2 is important, but less than the 52% of erroneous navigations observed in [10]. In these cases, physician decisions to follow the erroneous OncoDoc2 advices in 21.4% of the cases, and compliance with CPGs falls to 65.5%. This illustrates the unintended consequences of CDSSs and serves as a reminder that the accuracy of collected data, as well as CDSS usability are fundamental issues to address. Further analysis should investigate the determinants of Advice- situations.

The risk of Advice- situations is to favor the automation bias, and the compliance with CDSS, in the detriment of the compliance with CPGs. With our dataset, and according to the proposed formalization of users' attitudes, automation bias was measured at 4.6% of the CDSS-supported decisions. This

is lower than the 6 to 11% rates reported by the review conducted by Goddard et al. [18].

In 16.0% of the cases, MDM physicians positively reacted (React+) to the system. This demonstrates that they critically considered inappropriate CDSS advices and that in these cases, physicians' knowledge of CPGs is strongly established. On the opposite, negative reactance (React-) occurred when CDSS advices were not adopted and the decision was not CPG-compliant. This suggests that either CPGs are not known, or are not trusted, or are not appropriate to the patient. Indeed, in some particular clinical situations, deviating from CPGs can be either necessary since recommendations may not apply. This highlights the limits of CPG-based care and the room for clinical expertise, as promoted by evidence-based medicine.

This study presents some limitations. First, the reference navigation were built from the data found in patient records and considered as the gold standard, but might not be the best characterization of the patient. Errors in the patient record could lead to erroneous reference navigation. Thus an erroneous reference navigation is compared to an accurate MDM navigation. Some data input by MDM physicians considered as erroneous compared to the reference could in fact come from their clinical interpretation. For instance, they may consider that a patient could not undergo a chemotherapy, while nothing in the patient record suggests it. In addition, when the reference navigation is correct, physicians may have good clinical reasons not to comply with the CDSS advices for the benefit of the patient. This occurs for instance in the case of particular patient characteristics like BRCA1 mutation, patient preferences, or evolutions of the state of the art, where the propositions of the CDSS are not the best up-to-date recommendations for the patient. Automation bias may also be hidden behind Comp+ situations, where users follow the CDSS advices, and comply to CPGs while it would be better for a particular patient to deviate from CPGs. The same could be observed in React- situations. For instance, in the case of a surgery by mastectomy and axillary dissection is recommended, CDSS advice is tumorectomy, and decision is mastectomy, then, the decision neither complies with the CDSS nor with CPGs and is classified as React-. However, if the surgical treatment is split into two different surgical steps, mammary and axillary surgeries, we observe that there is a positive reactance (React+), from the advised tumorectomy to the recommended mastectomy for the mammary surgery, and a negative compliance (Comp-), since no axillary treatment is advised or decided albeit recommended. As a conclusion, the effect of the CDSS on the decision-making process is more complex than our proposed formalization, especially for complex multiple-step decisions. But these effects are hardly quantifiable since not detectable.

Conclusion

This study demonstrates that, when wrongly used, a guideline-based CDSS could deliver non guideline-based advices that hinder compliance of decisions with CPGs and favors user attitudes like negative reactance and automation bias. Hopefully, the advantages of HIT could outweigh the disadvantages. In our study, global compliance with CPGs was improved in the intervention arm, when OncoDoc2 was used [10]. The challenge is now to foster CPG compliance while warrantying patient safety through correct use of CDSSs. Potential solutions include improving CDSS user interface, ensuring data quality, and training physician to both CDSS and CPGs while adopting a critical appraisal on CDSS output.

Acknowledgments

Authors thank all the MSM members of the centers that participated in the study, the DSI/AP-HP that developed the OncoDoc2 system, and URCEST for their logistic support.

References

[1] Wennberg JE. Forty years of unwarranted variation—and still counting. Health Policy 2014;114(1):1–2.

[2] Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012;7:50.

[3] Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013;346:f657.

[4] Wöckel A, Kurzeder C, Geyer V, Novasphenny I, Wolters R, Wischnewsky M, et al. Effects of guideline adherence in primary breast cancer–a 5-year multi-center cohort study of 3976 patients. Breast 2010;19(2):120–7.

[5] van Hoeve J, de Munck L, Otter R, de Vries J, Siesling S. Quality improvement by implementing an integrated oncological care pathway for breast cancer patients. Breast 2014;23(4):364–70.

[6] Keating N, Landrum M, Lamont E, Bozeman S, Shulman L, McNeil B. Tumor boards and the quality of cancer care. J Natl Cancer Inst 2013;105(2):113–21.

[7] Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M. Cancer multidisciplinary team meetings: Evidence, challenges, and the role of clinical decision support technology. Int J Breast Cancer 2011:831605

[8] Séroussi B, Bouaud J, and Antoine ÉC., OncoDoc, a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. Artif Intell Med 2001;22 (1):43–64.

[9] Séroussi B, Laouénan C, Gligorov J, Uzan S, Mentré F, Bouaud J., Which breast cancer decisions remain non-compliant with guidelines despite the use of computerised decision support? Br J Cancer 2013;109(5):1147–56.

[10] Bouaud J, Blaszka-Jaulerry B, Zelek L, Spano JP, Lefranc JP, Cojean-Zelek I, et al. Health information technology: use it well, or don't! Findings from the use of a decision support system for breast cancer management. In: AMIA Annu Symp Proc. 2014:315–24.

[11] Campbell EM, Sittig D, Ash J, Guappone K, Dykstra R. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2006;13(5):547–56.

[12] Weiner J, Kfuri T, Chan K, Fowles J. "e-iatrogenesis": the most critical unintended consequence of CPOE and other HIT. J Am Med Inform Assoc 2007;14(3):387–8.

[13] Sittig D, Singh H. Defining health information technology-related errors: new developments since to err is human. Arch Intern Med 2011;171(14):1281–4.

[14] Magrabi F, Aarts J, Nohr C, Baker M, Harrison S, Pelayo S, et al. A comparative review of patient safety initiatives for national health information technology. Int J Med Inform 2013;82(5):e139–48.

[15] Vashitz G, Meyer J, Parmet Y, Peleg R, Goldfarb D, Porath A, et al. Defining and measuring physicians' responses to clinical reminders. J Biomed Inform 2009;42(2):317–26.

[16] Goddard K, Roudsari A, Wyatt JC. Automation bias: Empirical results assessing influencing factors. Int J Med Inform. 2014;83(5):368–75.

[17] Campbell EM, Sittig DF, Guappone KP, Dykstra RH, Ash JS. Overdependence on Technology: An Unintended Adverse Consequence of Computerized Provider Order Entry. In: AMIA Annu Symp Proc. 2007:94–8.

[18] Goddard K, Roudsari A, Wyatt JC. Automation bias—a systematic review of rates, mediators and mitigators. J Am Med Inform Assoc 2012;19(1):121–7.

Address for correspondence

Jacques Bouaud
LIMICS – INSERM U1142
15, rue de l'école de médecine, 75006 Paris, France
jacques.bouaud@sap.aphp.fr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-270

Use of a Proven Framework for Computer Decision Support within the Intermountain Healthcare Network

R. Scott Evans[ab], James F. Lloyd[a], Kyle V. Johnson[a], Stephen Howe[a], Jacob S. Tripp[a]

[a]*Department of Medical Informatics, Intermountain Healthcare*
[b]*Department of Biomedical Informatics, University of Utah*

Abstract

Hospitalized patients in the U.S. do not always receive optimal care. In light of this, Computerized Decision Support (CDS) has been recommended to for the improvement of patient care. A number of methodologies, standards, and frameworks have been developed to facilitate the development and interoperability of computerized clinical guidelines and CDS logic. In addition, Health Information Exchange using Service-Oriented Architecture holds some promise to help realize that goal. We have used a framework at Intermountain Healthcare that employs familiar programming languages and technology to develop over 40 CDS applications during the past 13 years, which clinicians are dependent on each day. This paper describes the framework, technology, and CDS application development methods, while providing three distinct examples of applications that illustrate the need and use of the framework for patient care improvement. The main limitation of this framework is its dependence on point-to-point interfaces to access patient data. We look forward to the use of validated and accessible Service-Oriented Architecture to facilitate patient data access across diverse databases.

Keywords:

Computerized Decision Support; CDS; framework; Service-Oriented Architecture; SOA.

Introduction

Hospitalized patients expect to receive the best possible care. However, of 25 clinical conditions examined, patients in the U.S. only received the recommended processes for basic care an average of 54.9% of the time [1]. Use of information technology has been recommended to help improve patient care [2, 3]. Computerized Decision Support (CDS) has been defined as "any computer program designed to help health professionals make clinical decisions" [4]. Airline pilots cannot look out the cockpit window and determine their exact altitude, air speed and flight direction. On-board computer programs provide that information. Likewise, if pilots try to land without the landing gear down and locked, or numerous other potential problems are detected during flight, the on-board computers generate alerts notifying the pilots of the situation and danger. In a similar manner, CDS has been used to provide real-time information, reminders and alerts to healthcare providers over the past 40 years, and has been shown to improve patient care [5, 6]. For more than 20 years, a number of methodologies, standards, intelligent systems, syntaxes and frameworks have been developed to facilitate the development and interoperability of computerized clinical guidelines and CDS logic. Protégé, GLIF, Arden Syntax, Prodigy, SEBASTIAN, GELLO and SAGE have been successful in demonstrating the reconstruction of previous CDS logic and the use of standards and ontologies to facilitate CDS and computerized guideline interoperability [7-14]. Health Information Exchange (HIE), using a Service-Oriented Architecture (SOA), provides some additional capacity to realize the goal of data independence [15-18]. Recently, the Substitutable Medical Applications, Reusable Technologies (SMART) project has recommended the development of substitutable applications built around core electronic medical records (EMR) [19]. To date, using the Fast Healthcare Interoperability Resources (FHIR) data format standards for web-based application programming interfaces (API), SMART looks like the best candidate for API development [19]. However, HIE efforts are not ready to be used and historically, committee acceptance has been difficult to obtain [20]. For over 13 years, we have used a framework at Intermountain Healthcare that uses familiar programming languages and technology to develop over 40 CDS applications, which clinicians depend on each day. This paper describes the framework, technology and CDS application development methods and provides three different examples of applications that illustrate the need for, and how we use the framework to improve patient care.

Materials and Methods

Background

Intermountain Healthcare (IH) spans the U.S. states of Utah and Idaho, and is comprised of 22 hospitals and 197 clinics, in addition to urgent care centers, physician offices and home health.. A key feature of our independently developed hospital information system (HELP) is the integrated EMR that contains most clinical information. The coded data in the EMR facilitates the development and use of CDS applications to analyze patient data and constantly monitor patient care. Natural Language Processing (NLP) is also used to analyze non-coded dictated reports.

In 1995, development on the HELP^2 system was started and ambulatory data in addition to hospital data began to be entered into the EMR. HELP^2 was expected to replace HELP before 2005. Moreover, during the past 20 years, IH has partnered with four different EMR vendors to develop a new information system, which would replace HELP and HELP^2. The HELP system and two vendor partners' EMRs were built on a hierarchical data dictionary while HELP^2 and the other two vendor partners' EMRs were built on relational databases and different data dictionary coding. Thus, for over 15 years,

the needs of IH clinicians for new and updated CDS applications continued to grow while we were not sure which EMR or data dictionary would eventually be used at IH nor which clinical guideline, medical logic development or HIE would become the standard.

Framework Description

Currently, the EMR at IH contains data from HELP and HELP^2 (Figure 1). Eventually, all the data in the EMR will only be provided by Cerner, the current partner, as HELP and HELP^2 are replaced at all IH hospitals and clinics. Each night, all clinical and business data in the EMR are archived in the enterprise data warehouse (EDW). After the EDW has been updated, Extract, Transform and Load (ETL) jobs are activated by the time driver and use SQL to extract and populate numerous Oracle tables on the staging server. The staging server is maintained 24/7 and backed up each night. The staging server provides faster data access as opposed to collecting that data from the large, slower and non-24/7 EDW. The data on the staging server are stored in data marts oriented by specific clinical needs of decision support applications. The "time driver" (Table 1) is also used to activate over 40 Java based CDS applications that run from once a month to every 5 seconds. The enterprise encounter table contains the master list of all inpatient and outpatient encounters along with some demographic data. The first thing most of the CDS applications do is look in the enterprise encounter file for patients that match specific requirements specified in the logic. The logic in each of the CDS applications also has a specific list of needed patient data elements within specific time windows. An application may require a variety of different patient data elements that could have been stored in any inpatient or outpatient facility from a number of years back in time or just during the current encounter. Historical data from other facilities is retrieved from the tables on the staging server, data for monthly reports are retrieved from the EDW, and current encounter data are retrieved from the EMR. A "data driver" monitors the EMR and is activated by data entry to the system. The data driver then runs specific medical logic modules based on the data being stored. If the medical logic module detects an action that needs to be made, an alert is stored in the Alert File on the IH network. Additionally, there exist data types not stored in the EMR. Electrocardiograms, echocardiograms, pulmonary function tests and data from the IH Health Plans are stored in different databases but also on the IH network.

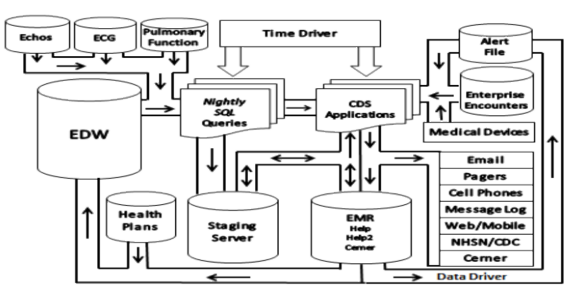


Figure 1 - Framework used for clinical decision support applications at Intermountain Healthcare. NHSN-National Healthcare Safety Network, CDC-Center for Diseases Control and Prevention.

Medical Devices Interface

Some CDS applications also need access to data from medical devices on the framework such as ventilators, infusion pumps, bedside and portable monitors, etc. Collection of data from medical devices and stored in tables on the staging server or the EMR is straightforward if the device has an interface that allows data to be transmitted to an external device. That interface can be analog or digital. If it is analog, the data must be converted from an analog signal to a digital interpretation of the signal. Many common analog-to-digital (A/D) converters are available to perform that function. If the data comes from the medical device in digital format, it may be in a number of different formats, e.g., binary or ASCII. Binary data are generally decoded to form some kind of human understandable format so that they can be stored as coded data. If the data are in ASCII, it may follow several standard protocols, e.g. HL7, XML, or it may be in a manufacturer's proprietary format, which may also need to be decoded to extract the desired components from the data. Hardware interfaces on the manufacturer's devices range from 2 wires for an analog output, to RS232 interfaces for simple serial data transmission and wired Ethernet connections, or a wireless interface including 802.11 (WiFi) or 802.15.4 (Zigbee).

Through the use of this framework, the logic in the CDS applications has access to all data within the IH network and not just a single hospital's EMR. Some data stored in the staging server are not needed until a patient is re-admitted or scheduled for a clinic or home health visit the next day. Actions from the CDS applications can be stored in the EMR or include alerts that can be sent to email addresses, pagers/cell phones, the Message Log within the HELP^2 EMR, Web and mobile devices, the National Healthcare Safety Network at the Centers for Diseases Control and Prevention, or through an API or service to a vendor. Data in the EMR from HELP^2 is accessed through Oracle SQL queries imbedded in Java code while HELP data are extracted using a client/server system of components developed in-house.

CDS Application Development

The CDS applications we have developed on the framework range from fairly simple alerts such as sending primary care physicians daily alerts in the HELP^2 "Message Log" stating that ECG results are finalized to a complex web-based version of the Antibiotic Assistant program [21]. We have found Java to work well and provide us with all the methods we need for simple to complex CDS logic (Table 1). We use Eclipse as the integrated development environment (IDE) for the CDS application development (Eclipse Foundation, Inc., Ottawa, Ontario, Canada). In addition to being free and open source, it provides all the development and testing tools we need; source code editor, intelligent code completion, build automation, documentation generator and especially debugging tools. For complex applications, we need a symbolic debugger that provides the ability to step through code, set break points, break when variables are changed, modify variables and rerun the program from specific lines of code. SQL embedded within the Java code is used to extract the needed data elements from relational databases through Java database connectivity. Java classes also handle the HELP data extraction interface or point-to-point interfaces for the electrocardiogram, echocardiogram and pulmonary function databases and medical devices.

We use JBoss middleware to write to queues and buffer data sent to the EMR. While we use in-house developed services to display alerts to our HELP^2 Message Log, we use common methods to provide most alerts and reports to the users. Simple mail transfer protocol (SMTP) is used to send alerts to

pagers and cell phones and lengthy text reports are sent via email. Reports are encoded using XML to send monthly antibiotic use data to the Center for Diseases Control and Prevention. JavaScript is also used to display CDS derived output on the web.

Table 1 – List of CDS tasks and common methods used within the framework.

CDS Task	Common Method
Core programming language	Java
Integrated development environment: source code editor, intelligent code completion, build automation, documentation generator, symbolic debugger	Eclipse
Time driver	Windows Scheduler
Relational data access	SQL embedded within Java and Java Database Connectivity
Non-relational data access	Point-to-point developed Java classes
Messages within the HELP[2] EMR	In-house developed Java classes
Pager, cell phone and email alerts	Simple Mail Transfer Protocol
Medical device access and data storage	RS232 interfaces, analog-to-digital converters, HL7, XML, wifi, JBoss
Antimicrobial use reports to Center for Diseases and Control	XML
Web-based report generation	JavaScript and HTML within Java

Application Examples

Ventilator disconnection: In 2004, we developed a system to monitor critical ventilator events by using the framework [22]. We collect data from mechanical ventilators every 5 seconds and look for evidence of a disconnection. Whenever an event is identified, the system takes control of every computer in the patient's intensive care unit and generates an enhanced audio and visual alert indicating there is a critical ventilator event and identifies the room number. Once the alert is acknowledged or the event is corrected, all the computers are restored back to the pre-alert status and/or application state. That CDS application was first installed in the Shock/trauma intensive care unit (ICU) at LDS Hospital. Today, the ventilator disconnection alerts are installed in 13 ICUs at 6 different IH hospitals. The audio and video alerts are so distinct and annoying that all medical personnel quickly respond to get the alerts turned off. Data from the past 10 years show that the present critical ventilator alarm times are below durations that are likely to be dangerous. In addition, now that data is collected on all ventilator disconnections, our respiratory therapy departments have been able to identify risk factors such as where ventilators are placed in relation to heat and air conditioning vents which can cause water accumulation in the tubing and restrict airflow. This CDS example illustrates the need for data access with the encounter data, medical devices, EMR, the time driver and use of an IDE for complex code development.

Venous thromboembolism (VTE) high-risk alerts: Routine use of evidence-based guidelines for VTE prophylaxis is uncommon due to the difficulty in integrating them into routine patient care**.** We developed a CDS tool which captures sufficient data to use a risk prediction model to identify patients at high risk for VTE [23]. This CDS application illustrates the need for encounter, historical and current patient data from throughout the enterprise. Patients are screened each day of hospital admission at all 22 IH hospitals and their risk of VTE is determined. Pharmacists are sent an email message listing each high-risk patient and which risk factors they have. Patients at high-risk are then checked to see if they are on appropriate VTE chemoprophylaxis including the drug, dosage, route and interval. If not, 65 hospitalists get a page if their patient is identified. Patient age, new surgery, bed-rest, cancer, height and weight information are collected from the EMR. Previous cancer, previous VTE, previous surgery within 90 days, hypercoagulability and hormone replacement and oral contraception use are collected each night from the EDW and stored in the staging server. The time driver activates the application at 7am each day. We found this computerized tool to identify patients at high risk for VTE with a sensitivity of 98% and positive predictive value of 99%. As of January 2014, high and low risk alerts are sent for every patient to the nurse-charting program in the HELP EMR to meet the new Center for Medicare and Medicaid Service mandate in the U.S.

Antibiotic Assistant: The most complex CDS application we have developed is the antibiotic assistant [21]. Most of the data are accessed from the EMR except for previous microbiology, laboratory, radiology and pathology data from other hospitals during the previous 6 weeks. The application displays all the information physicians need to be aware of before they order antimicrobial agents or are checking the infection status of their patients. The most complex part is the suggestion of which antimicrobial(s) along with the dosage, route, interval and duration to use. The application was originally developed using Tandem Application Language and a symbolic debugger on HELP. With the eventual sunsetting of HELP, we were able to duplicate the application using Java and JavaScript so it can run on HELP[2] or any other system that can render web-based applications such as most vendor EMRs. Based on the complexity of the logic and number of interdependent Java classes in this application, it is used to illustrate the need for a symbolic debugger. It would not have been possible to develop, maintain or add new features to this CDS application without the functionality provided by a symbolic debugger - an ability provided by both Java and Eclipse.

Discussion

The CDS framework and application development methods described in this paper have served us well for the past 13 years, and have allowed us to meet the diverse information needs of many different types of clinicians and medical domains. The majority of CDS applications that we have developed are a) time-driven, b) run in the background, and c) send alerts or email reports at programed intervals in time designated by clinician workflow and patient care. These applications are well aligned with the four predictors of improved clinical practice: 1) automatic provision of decision support as part of clinician workflow, 2) provision of recommendations rather than just assessments, 3) provision of decision support at the time and location of decision making, and 4) computer based decision support [6].

At present, there is no widely accepted, standard approach for representing CDS intervention types and their structural components including alerts and reminders, order sets, infobuttons, documentation templates/forms, and relevant data presentation [24]. Many gaps and challenges remain, including the complexity of many standards, limited availability of accessible knowledge resources, and lack of tooling and other resources to enable the adoption of existing standards [25]. Moreover, many question the "outsourcing" of CDS to potentially imperfect external agencies [26].

While system-agnostic CDS development tools hold great potential to facilitate CDS implementation, little has been described regarding their benefits and challenges (26). A benefit of using Java and SQL is that they are common programming languages taught within most computer science departments. Thus, no other special knowledge or training is required to learn how to use a new CDS development tool and syntax by the programmers. Our experience is that most clinicians are extremely busy providing and looking for ways to improve patient care and don't have the time for or an interest in learning how to program CDS applications. Also, use of Java classes and libraries provide an easy way to share and reuse the same developed and tested logic from one application to another. While not quite true, the Java write once, run anywhere java virtual machine does facilitate the transfer of CDS applications to many different platforms. We do not imply that this framework ought to replace others, nor do we predict that it will outperform any of the other CDS development methods available. Rather, we contribute this framework as a viable option for other systems, given its high functionality in application development and reduction of testing time at IH.
Use of this framework has taught us that the ability of CDS to improve patient care depends on reliable access to the needed patient data. Often, those data need to be derived in real time from the EMR or medical devices, and not the EDW. We are fortunate to have access to most patient data using this framework we have built at IH. Other institutions can use these same programming and development methods we describe to create new CDS applications using the data to which they have access.

The main constraint of our framework is its limiting accessibility to EHR patient data, by way of medical device and database point-to-point interface. Accessing patient data across multiple EHRs has been a desired but largely unattained aim of clinical informatics, especially in commercial EHR systems (24). A potential opportunity for enabling CDS across multiple EHRs is to leverage vendor-supported SOA development and APIs. A recent report by JASON, an independent group of scientists who advise the U.S. federal government on technology issues, states that to achieve the goal of improving health outcomes, Stage 3 Meaningful Use requirements should be defined such that they enable the creation of an entrepreneurial space across the entire health data enterprise and suggest that EHR software vendors should be required to develop and publish APIs for EMRs for third-party software developers (http://healthit.gov/sites/default/files/ptp-13700hhs_white.pdf).

SMART's approach is to use FHIR, a HL7 draft standard describing data formats and elements and an API, to help drive down healthcare costs, support standards, accommodate differences in workflow, and especially foster competition in the market to help promote "innovation" in healthcare [19]. To date, we have used services provided by Cerner to access data for two of the clinical applications that run on our framework. Cerner's upper management has publically stated they support the need for vendors to provide SOA and especially use SMART on FHIR. If so, when a SOA is finalized and in use, we would be able to replace the current point-to-point database specific services and continue to be innovative without the need to make any changes to our current framework, technology or CDS logic.

Conclusion

We have used the described framework, which has employed familiar programming languages and technology in the development over 40 clinician dependent CDS applications during the past 13 years. The main limitation of this framework is the dependence on point-to-point interfaces to access patient data. We look forward to when we can use validated and accessible SOA to facilitate patient data access across diverse databases.

References

[1] McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. The New England journal of medicine. 2003 Jun 26;348(26):2635-45.

[2] Medicine Io. Crossing the quality Chasm: a new health system for the 21st century. National Academy Press. 2001;Washington DC

[3] Executive Office of the President PsCoAoS, technology. a. Report to the President: Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: the Path Forward. 2010:http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-itreportpdf.

[4] Shortliffe EH. Computer programs to support clinical decision making. Jama. 1987 Jul 3;258(1):61-6.

[5] Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. Jama. 1998 Oct 21;280(15):1339-46.

[6] Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj. 2005 Apr 2;330(7494):765.

[7] Musen MA, Gennari JH, Wong WW. A rational reconstruction of INTERNIST-I using PROTEGE-II. Proceedings / the Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care. 1995:289-93.

[8] Peleg M, Boxwala AA, Bernstam E, Tu S, Greenes RA, Shortliffe EH. Sharable representation of clinical guidelines in GLIF: relationship to the Arden Syntax. Journal of biomedical informatics. 2001 Jun;34(3):170-81.

[9] Pryor TA, Hripcsak G. The Arden syntax for medical logic modules. International journal of clinical monitoring and computing. 1993 Nov;10(4):215-24.

[10] Johnson PD, Tu S, Booth N, Sugden B, Purves IN. Using scenarios in chronic disease management guidelines for primary care. Proceedings / AMIA Annual Symposium AMIA Symposium. 2000:389-93.

[11] Borbolla D, Otero C, Lobach DF, Kawamoto K, Gomez Saldano AM, Staccia G, et al. Implementation of a clinical decision support system using a service model: results of a feasibility study. Studies in health technology and informatics. 2010;160(Pt 2):816-20.

[12] Jenders RA, Del Fiol G, Kawamoto K, Sailors RM. Standards in clinical decision support: activities in health level seven. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2008:1244-5.

[13] Park JY, Musen MA. VM-in-Protege: a study of software reuse. Studies in health technology and informatics. 1998;52 Pt 1:644-8.

[14] Tu SW, Musen MA, Shankar R, Campbell J, Hrabak K, McClay J, et al. Modeling guidelines for integration into clinical workflow. Studies in health technology and informatics. 2004;107(Pt 1):174-8.

[15] Loya SR, Kawamoto K, Chatwin C, Huser V. Service oriented architecture for clinical decision support: a systematic review and future directions. Journal of medical systems. 2014 Dec;38(12):140.

[16] Kawamoto K, Jacobs J, Welch BM, Huser V, Paterno MD, Del Fiol G, et al. Clinical information system services and capabilities desired for scalable, standards-based, service-oriented decision support: consensus assessment of the Health Level 7 clinical decision support Work Group. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:446-55.

[17] Kawamoto K, Hongsermeier T, Wright A, Lewis J, Bell DS, Middleton B. Key principles for a national clinical decision support knowledge sharing framework: synthesis of insights from leading subject matter experts. Journal of the American Medical Informatics Association : JAMIA. 2013 Jan 1;20(1):199-207.

[18] Kawamoto K, Lobach DF. Proposal for fulfilling strategic objectives of the U.S. Roadmap for national action on clinical decision support through a service-oriented architecture leveraging HL7 services. Journal of the American Medical Informatics Association : JAMIA. 2007 Mar-Apr;14(2):146-55.

[19] Mandl KD, Mandel JC, Murphy SN, Bernstam EV, Ramoni RL, Kreda DA, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. Journal of the American Medical Informatics Association : JAMIA. 2012 Jul-Aug;19(4):597-603.

[20] Mandl KD, Kohane IS. No small change for the health information economy. The New England journal of medicine. 2009 Mar 26;360(13):1278-81.

[21] Evans RS, Pestotnik SL, Classen DC, Clemmer TP, Weaver LK, Orme JF, Jr., et al. A computer-assisted management program for antibiotics and other antiinfective agents. The New England journal of medicine. 1998 Jan 22;338(4):232-8.

[22] Evans RS, Johnson KV, Flint VB, Kinder T, Lyon CR, Hawley WL, et al. Enhanced notification of critical ventilator events. Journal of the American Medical Informatics Association : JAMIA. 2005 Nov-Dec;12(6):589-95.

[23] Evans RS, Lloyd JF, Aston VT, Woller SC, Tripp JS, Elliott CG, et al. Computer surveillance of patients at high risk for and with venous thromboembolism. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2010;2010:217-21.

[24] Zhou L, Hongsermeier T, Boxwala A, Lewis J, Kawamoto K, Maviglia S, et al. Structured representation for core elements of common clinical decision support interventions to facilitate knowledge sharing. Studies in health technology and informatics. 2013;192:195-9.

[25] Kawamoto K, Del Fiol G, Lobach DF, Jenders RA. Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress. The open medical informatics journal. 2010;4:235-44.

[26] Nadkarni PM, Miller RA. Service-oriented architecture in medical software: promises and perils. Journal of the American Medical Informatics Association : JAMIA. 2007 Mar-Apr;14(2):244-6.

Address for correspondence
R. Scott Evans, M.S., Ph.D., FACMI
Dept. Medical Informatics, LDS Hospital
8th Avenue & C Street
Salt Lake City, Utah, 84143 USA.
rscott.evans@imail.org (801) 408-3029

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-275

An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders

Alexandre Galopin[a,b,c,d], Jacques Bouaud[e,b,c,d], Suzanne Pereira[a], Brigitte Seroussi[c,f,b,d]

[a] *Vidal, Issy-les-Moulineaux, France*
[b] *INSERM, U1142, LIMICS, Paris, France*
[c] *Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, LIMICS, Paris, France*
[d] *Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), Bobigny, France*
[e] *AP-HP, DRCD, Paris, France*
[f] *AP-HP, Hôpital Tenon, DSP, Paris, France; APREC, Paris, France*

Abstract

Decision support systems, as means of disseminating clinical practice guidelines, are powerful software that may lead to an improvement of medical practices. However, they are not always efficient and may suffer from limitations among which are lack of flexibility and weaknesses in the integration of several clinical practice guidelines (CPGs) for the management of patients with multiple chronic disorders. We propose a framework based on an ontological modeling of CPG contents as rules. The ontology provides the required flexibility to adapt patient data and enable the provision of appropriate recommendations expressed at various levels of abstraction. To solve decisional conflicts that occur when combining multiple sources of recommendations, we proposed a method based on the subsumption graph of the patient profiles corresponding to the rules. A prototype CDSS implementing this approach has been developed. Results are given on a clinical case to illustrate the assets of ontological reasoning in increasing the number of issued recommendations and thereby the reliability of decision support.

Keywords:

Practice guidelines as topic; Knowledge bases; Ontology; Clinical decision support systems; Medication therapy management.

Introduction

Delivering optimal quality of care for all is one of the major challenges of modern medicine. Unwarranted variations of medical practices should be reduced [1], and improvement of care based on solid evidence should be promoted. Health agencies and medical professional societies develop clinical practice guidelines (CPGs), which are textual documents that synthesize the state of the art on the management of medical disorders. CPGs are expected to provide decision support for certain clinical situations, being one instrument of the promotion of evidence-based medicine. So far, their implementation remains insufficient. The challenge consists in a clear understanding of the reasons that prevent the use of CPGs [2], and the development of interventions to increase their implementation [3].

Clinical decision support systems (CDSSs) are considered as an efficient vector of CPG implementation. CDSSs rely on formalized knowledge bases and are usually integrated to electronic patient records to assist the clinician in her everyday practice. The integration of CPGs in CDSSs requires a preliminary formalization step of CPG contents [4]. CPG knowledge may be represented as a set of decision rules where each rule consists in the description of a given patient profile and its associated recommendation. Therefore, CPGs are modeled as clinical patient profiles for which recommendations are provided. Guideline-based CDSSs yield patient-specific recommendations by pairing the description of the actual patient with related CPG patient profiles. Despite their promise, several studies reported discrepancies between CDSS expectations and their effectiveness in promoting best practices, illustrating the existence of obstacles to their widespread use in routine [5, 6].

CPGs generally focus on a specific medical disorder, eg. Hypertension, Asthma, Obesity, etc. But, actual patients often present multiple pathologies and the management of multi-morbidity can be a real challenge for the clinician [7]. Indeed, this requires identifying CPGs related to the patient state, to gather every relevant recommendation, and to combine them accordingly [8]. Competing CPGs and their potential decisional conflicts are identified as a reason of non-adherence with CPGs [2]. Future CDSSs should account for the integration of multiple CPGs [9].

Clinical descriptions of theoretical patient situations are often not detailed in the same way depending on CPGs. Indeed, CPGs are usually written in free text by different consortia, sometimes in different languages. Therefore, the same concept might be described using different terms, and with stylistic variations. A standardized description is then required for representing CPG knowledge. Depending on CPGs, some patient characteristics may be described at different levels of abstraction, which requires the tacit knowledge that link them. For instance, kidney disease is mentionned as a co-morbidity in diabetes CPGs, whereas more specific precisions in terms of renal failure are considered in hypertension CPGs. This lack of normalization makes the identification of appropriate patient states complex. It may lead to an incomplete or bad characterization of the patient and consequently to the production of conflicting recommendations, incorrect recommendations, or even no recommendation. Indeed, it happens that guideline-based CDSSs are not able to match patient data with the premisses of recommendations, which can result in a problematic silence when the system does not provide any recommendation to the user. This may cause a loss of confidence and usually ends up with an abandonment of such systems.

While developping biomedical applications, ontologies (domain conceptualizations for representing CPG knowledge) can efficiently be used as a stable framework [10]. Our

assumption is that they could provide solutions to the seamless consideration of multiple CPGs within CDSSs.

We have developed a method based on ontologies to give CDSSs the flexibility needed to deal with patients with multiple pathologies. The aim of this paper is to describe the framework we developed where ontological reasoning is used to enrich the patient description at different levels of abstraction and thereby increase the number of appropriate recommendations. The approach allows to solve decisional conflicts exploiting the subsumption of CPG-based patient profiles in order to provide the most relevant recommendations. We have implemented a prototype CDSS applied to CPGs on hypertension (HT) and type 2 diabetes (T2D). We used an example to illustrate the impact, on the quality of the recommendations provided, of taking into account the ontological reasoning.

Methods

Our goal is to address the management of multiple CPGs, expressed at various levels of abstraction, within a functional CDSS. The approach relies on the use of ontological representation and reasoning. The first step deals with the construction of the knowledge base from multiple CPGs. The second focuses on the exploitation of the knowledge base and the processing of patient data to deliver patient-specific recommendations. The proposed framework aims at increasing the number of patient-specific recommendations while managing potential conflicts.

Hypertension and type 2 diabetes CPGs

Considering the management of cardiovascular risk in general practice, we first chose to consider HT and T2D. We used contemporary CPGs authored in 2014 by Vidal, a French company that markets a drug database and original medical content whose quality has been certified by the medical profession. CPGs include graphical clinical pathways that illustrate the process of care. Each step of the trees are then detailed by synthetic textual sections. CPGs also deal with particular cases (eg. HT and Pregnancy) and provide indications and contraindications of recommended drug classes.

Construction of the knowledge base

In a previous study [11], we performed a formalization of HT and T2D CPGs by manually extracting and conceptualizing the decision rules. As illustrated in Figure 1, decision rules are built on the IF-THEN model with the conditional combination of patient criteria in the IF-part and the recommended actions in the THEN-part. The IF-part corresponds to a patient profile described in CPGs. We obtained two rule bases: one for the HT CPGs and one for the T2D CPGs.

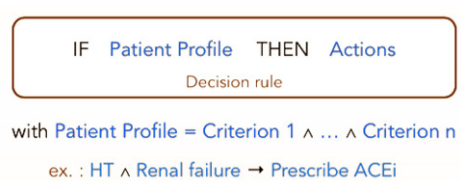


Figure 1 - Formalisation of decision rules

Decision criteria and actions were encoded in a single custom ontology. This ontology was inspired from the existing OntoIUrgences developed for the management of emergencies [12]. Concepts relevant to cardiovascular risk management were extracted from this emergency ontology, and supplemented by CPG-specific concepts required by our decision rules. We encoded the ontology in OWL. Concepts are related by subsumption, equivalence, and disjunction links so that every inference is valid. Figure 2 illustrates an extract of the custom ontology under the "Pathology" concept.

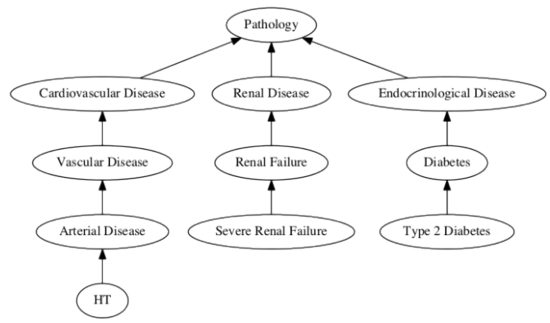


Figure 2 - Extract of the Pathology subgraph of the custom ontology

Generation of the graph of patient profiles

The patient profiles described in the IF-part of the rules are conjunctions of concepts from the ontology. Patient profiles correspond to new defined concepts and are consequently linked within the subsumption hierarchy of the ontology, some being more specific than the others. Following the ontological framework, there are two kinds of specificity: (i) the conceptual specificity which comes from the subsumption of atomic concepts (eg. "HT" is more specific than "Arterial Disease") and (ii) the logical specificity derived from defined concepts (eg. "HT ∧ Diabetes" is more specific than "HT"). The classification of patient profile/rules is performed automatically by an ontological reasoner. This yields a subsumption graph of patient profiles where the root is the least specified profile and the leaves are the most detailed ones. Figure 3 illustrates an extract of the profile subsumption graph for patient profiles of both HT and T2D rule bases.

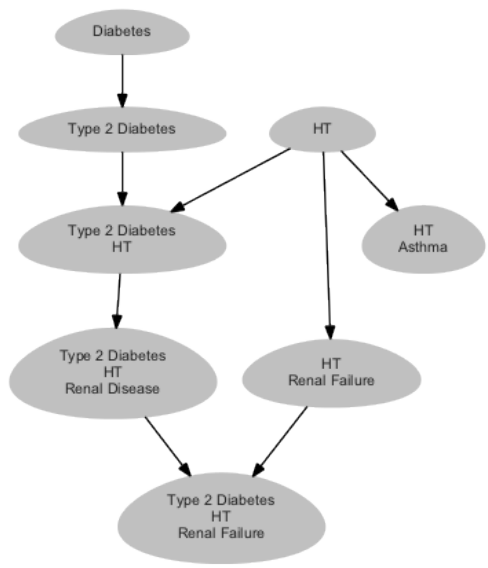


Figure 3 - Extract of the subsumption graph of profiles from HT and Type 2 Diabetes CPGs

Matching of decision rules

When an actual patient's clinical case is considered, patient data is first translated to match the concepts of the ontology. The actual patient description is then a conjunction of ontological criteria. The initial description of the patient is then enriched through the ontology. For instance, a patient described by "HT ∧ Diabetes" will be characterized by "HT ∧ Arterial Disease ∧ Vascular Disease ∧ Cardiovascular Disease ∧ Endocrinological Disease ∧ Pathology". In this framework, finding which decision rules apply for the patient consists in identifying all the CPG-based patient profiles that subsume the actual patient profile.

Management of potential conflicts

By using the subsumption of patient profiles, more rules than those that solely match patient data are, logically, considered for execution, whatever their level of abstraction, and whatever their originating CPG in case multiple rule bases from different CPGs have been merged. Potential conflicts may then be revealed and should be managed. We used the ontological representation of rule actions to detect such conflicts. When comparing every pair of matching profiles, we identified three possible cases:

1. No action conflict. Recommended actions are complementary and can be suggested without causing any decisional conflict. For instance, "Prescribe Anti-HT treatment" and "Monitor hypoglycemia" are two recommended actions of different nature and can thus be suggested together to the user.
2. Action conflict and profile subsumption. Recommended actions are conflicting and subsumption exists between the profiles from which actions have been extracted. The priority is given to the recommendation attached to the most specific profile. For instance, betablockers are recommended for patients suffering from HT, but contraindicated for patients suffering from HT and Asthma. Thus, the contraindication of betablockers overrides their recommendation in the case of a patient matching the "HT ∧ Asthma" profile.
3. Action conflict but no profile subsumption. In this case, recommended actions are conflicting but the originating profiles cannot be compared since there is no sumsumption link between them. Without additional encoded knowledge to solve this conflict, we cannot solve it automatically, and we chose to let the clinician decide what recommendation (if any) is the most relevant for his patient. For instance, for hypertensive patients, thiazide diuretics are recommended for diabetic patients but contraindicated for patient with renal failure. In the case of a patient characterized by "HT ∧ Diabetes ∧ Renal Failure", the system is unable to discriminate between the recommendations.

Analysis of the role of ontological reasoning

To assess the benefits of the ontological reasoning when merging several CPGs, we compared the results provided by the CDSS with the results of the same system when the ontological reasoning was disabled. To illustrate this, we used a simulated patient case profile which has to be a realistic situation of a patient suffering from multiple pathologies among which at least HT or Diabetes. Then, we executed the system and compared the characterization of the patient including inferred findings, triggered decision rules, and the recommendations of drug prescription.

Results

The custom CPG ontology built for HT and T2D CPGs is made of 500 concepts and includes 96 disjunction declarations. It is divided into two main parts: concepts related to the patient or decision variables (i.e. characteristics, pathologies, clinical signs) and those related to the clinical management or actions (i.e. treatment, goals, medical actions).

The rule bases contain 180 different rules for HT CPGs, and 94 different rules for T2D CPGs. These two rule bases share two common profiles ("Diabetes" and "HT ∧ T2D") which yields 272 different patient profiles for the unified rule base.

A CDSS has been developed to provide recommendations issued from HT and T2D CPGs. We used the JENA API for ontological reasoning, and rule inferences based on the set of patient data. A graphical user interface enables interaction with the user and the display of the recommended drug classes for each pathology as well as other recommended information (diet, prevention). Figure 4 displays a screenshot of the CDSS interface.

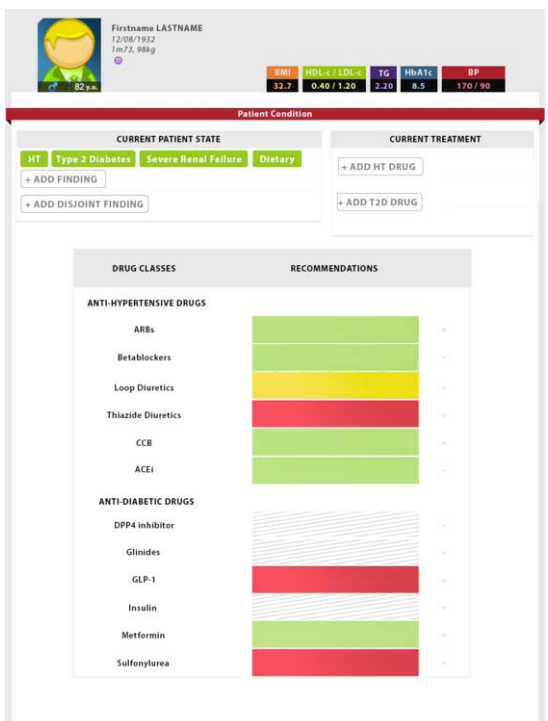


Figure 4 - Screenshot of the CDSS user interface

To illustrate how the system operates, we used the clinical case of a 83 year old male patient under dietary, with HT, T2D, and suffering from severe renal failure. His arterial pressure is 170/90 mmHg and his HbA1c rate is 8.5%. We run the system on this example patient with and without the ontological reasoning. Table 1 synthesizes the results obtained.

Without ontological reasoning, the system missed a contraindication to thiazide diuretics for people suffering from renal failure and recommended this class of antihypertensive drugs. When the ontological reasoning was enabled, the system detected a conflict about thiazide diuretics which are recommended for T2D patients, but are contraindicated for patient suffering from renal failure. Likewise, without ontological reasoning, the system missed the indication of

loop diuretics for patients with renal failure. Other actions that differ between the two execution modes of the system are not related to drug prescription and concern the other aspects of the management (risk factors, medical appointments, monitoring). It must be noted that the recommended drug prescription for the management of T2D is the same regardless of the activation of ontological reasoning.

Table 1 - Comparison of the CDSS execution with and without ontological reasoning on the example patient

Object	Without ontology	With ontological reasoning
Patient Concepts (input/inferred)	3/0 (HT, T2D, Severe Renal Failure)	3/11 (HT, T2D, Severe Renal Failure, *Arterial Disease, Vascular Disease, Cardiovascular Disease, Diabetes, Endocrinological Disease, Renal Failure, Renal Disease, Pathology*)
Incompatible Concepts	0	33
Incompatible profiles	0	136
Triggered profiles	15	22
Recommended actions	42	51
Contraindicated Anti-HT classes	—	Thiazide diuretics
Recommended Anti-HT classes	ARB, ACEi, Betablockers, Calcium-channel blockers, *Thiazide diuretics*	ARB, ACEi, Betablockers, Calcium-channel blockers, *Loop diuretics*
Recommended Anti-Diabetic classes	Metformin	Metformin
Contraindicated Anti-Diabetic classes	GLP-1, Sulfonylurea	GLP-1, Sulfonylurea

On the CDSS interface (see Figure 4), the drug prescription dashboard synthesizes the recommendations about drug prescription for the two pathologies. Drug classes are written on the left of the dashboard and the associated recommendations are indicated using color codes: red when the drug class is contraindicated, green when it is recommended, yellow when it is possible, and grey when no indication is given for this drug class in CPGs.

Discussion

In this paper we presented our work concerning the integration of ontological reasoning to handle multiple CPGs represented as decisional rules. The method has been implemented as a functional CDSS. We have illustrated the functionning of the system on an example clinical case. We compared the results obtained with and without ontological reasoning.

The number of concepts implicitely added by ontological reasoning is closely related to the construction of the ontology; it depends on the number of levels of specificity chosen during the modeling step. This choice has to be made by taking into account the levels of specificity of the decision rules. Indeed, a high number of additional concepts only has a meaning if it allows the triggering of more rules.

The CDSS with ontological reasoning provides, as expected, more recommendations than the classical approach. These additional recommendations were issued because rules of higher level of abstraction have been triggered. Decision rules used in the CDSS were indeed extracted from the text of CPGs and thus, inherited from text imperfections and inconsistency. Including ontological reasoning allowed to depart from the syntactical constraints of the text and to reduce such imperfections. By giving some flexibility in the characterization of a patient, we increased the number of triggered profiles including those that would not have been triggered by classical systems. The resulting recommendations are not only more numerous but also, thanks to our conflicts solving method, more adapted to the current patient case.

The lack of flexibility is pointed out as one of the reasons that clinicians don't keep using CDSSs. Ontologies bring the possibility to enrich the levels of abstraction in the management of patient profiles, and also to create, from static guidelines, dynamic pathways following the needs, the availability of the data and the usefulness of the recommendations. Other research works have incorporated ontologies for decision support as well as for other type of biomedical applications [13,14].

Besides content issues which are fundamental, but also apply to CPGs themselves, the usability of CDSSs and the clarity of the user interface are recognized as key factors for their adoption by health professionals. The real challenge lies in dealing with a lot of information and choosing which CDSSs to suggest to the clinician. We thus have to find a synthetic way to display the relevant provided recommendations and to highlight the most interesting ones. These aspects have to be further evaluated.

Conclusion

We propose a method based on semantic web techniques such as ontological reasoning to bring more flexibility to CDSSs and also offer the ability to deal with patients suffering from multiple pathologies by including several modeled CPGs. The method has been implemented as a prototype CDSS. In future works, we will enrich the current system with additional CPGs on the management of cardiovascular diseases (*e.g.* dyslipidemia). Then, we plan to assess the system both on the usability dimension and on the quality of the recommendations provided with a panel of general practictioners.

References

[1] Wennberg JE. Forty years of unwarranted variation—and still counting. Health Policy 2014: 114(1): 1–2.

[2] Flottorp S, Oxman A, Krause J, Musila N, Wensing M, Godycki-Cwirko M, et al., A checklist for identifying determinants of practice: A systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci 2013: 8(1): 35.

[3] Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, and Squires JE. Knowledge translation of research findings. Implement Sci 2012: 7: 50.

[4] Shiffman RN, Michel G, Essaihi A, and Thornquist E, Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. J Am Med Inform Assoc 2004: 11(5): 418–26.

[5] Jaspers MW, Smeulers M, Vermeulen H, and Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011: 18(3): 327–34.

[6] Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013: 346: f657.

[7] Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, and Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012: 380: 37-43.

[8] Hughes LD, McMurdo ME, and Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age Ageing 2013: 42(1): 62-9.

[9] Wilk S, Michalowski W, Michalowski M, Farion K, Hing MM, and Mohapatra S. Mitigation of adverse interactions in pairs of clinical practice guidelines using constraint logic programming. J Biomed Inform 2013: 46(2): 341-53.

[10] Schulz S, and Jansen L. Formal ontologies in biomedical knowledge representation. Yearb Med Inform 2013: 8(1): 132-46.

[11] Galopin A, Bouaud J, Pereira S, and Séroussi B. Using an Ontological Modeling to Evaluate the Consistency of Clinical Practice Guidelines: Application to the Comparison of Three Guidelines on the Management of Adult Hypertension. Stud Health Technol Inform 2014: 205: 38-42.

[12] Charlet J, Declerck G, Dhombres F, Miroux P, Gayet P, and Vandenbussche PY. Building a medical ontology to support information retrieval: terminological and metamodelization issues. Proc. of the10th International Conference on Terminology and Artificial Intelligence 2013: Oct 28-30: 171–8.

[13] Ramzan A, Wang H, and Buckingham C. Representing Human Expertise by the OWL Web Ontology Language to Support Knowledge Engineering in Decision Support Systems. Stud Health Technol Inform 2014: 207: 290-9.

[14] Riaño D, Real F, López-Vallverdú JA, Campana F, Ercolani S, Mecocci P, Annicchiarico R, and Caltagirone C. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform 2012: 45(3): 429-46.

Address for correspondence

Alexandre Galopin
VIDAL
21, rue Camille Desmoulins
92789 Issy Les Moulineaux Cedex 9, France
alexandre.galopin@vidal.fr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-280

Understanding Deviations from Clinical Practice Guidelines in Adult Soft Tissue Sarcoma

Esther Goldbraich[a], Zeev Waks[a], Ariel Farkash[a], Marco Monti[b,c], Michele Torresani[d], Rossella Bertulli[d], Paolo Giovanni Casali[d], Boaz Carmeli[a]

[a] *IBM Research, Haifa, Israel*
[b] *IBM, Milan, Italy*
[c] *Max Planck Institute for Human Development, Berlin, Germany*
[d] *Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy*

Abstract

In recent years we have witnessed the increasing adoption of clinical practice guidelines (CPGs) as decision support tools that guide medical treatment. As CPGs gain popularity, it has become evident that physicians frequently deviate from CPG recommendations, both erroneously and due to sound medical rationale. In this study we developed a methodology to computationally identify these deviation cases and understand their movitation. This was achieved using an integrated approach consisting of natural language processing, data modeling, and comparison methods to characterize deviations from CPG recommendations for 1431 adult soft tissue sarcoma patients. The results show that 48.9% of patient treatment programs deviate from CPG recommendations, with the largest deviation type being overtreatment, followed by differences in drug treatments. Interestingly, we identified over a dozen potential reasons for these deviations, with those directly related to the patients' cancer status being most abundant. These findings can be used to modify CPGs, increase adherence to CPG recommendations, reduce treatment cost, and potentially impact sarcoma care. Our approach can be applied to additional diseases that are subject to high deviation levels from CPGs.

Keywords:

Physician's Practice Patterns [N04.590.748],
Practice Guideline [V02.515.500],
Sarcoma [C04.557.450.795],
Decision Support Techniques [L01.700.508.190],
Natural Language Processing [L01.224.065.580].

Introduction

The modern medical landscape is characterized by a plethora of different treatment options for almost indistinguishable clinical statuses. While the development of new treatment modalities is beneficial, it also poses challenges associated with the growing body of evidence regarding the outcomes of different treatments.

As a consequence of the complexity of treatment possibilities and the presence of widespread variation in medical practice, it has become clear that a large fraction of patients do not in fact receive the best possible care [1,2]. Deviations from optimal care are abundant in diseases where treatment efficacy varies as a result of subtle changes in the clinical scenario as well as in cases where clear scientific evidence is not present, as is often seen in cancer [3,4]. Therefore, an important question in medicine is what leads clinicians to prescribe treatments that do not adhere to best practice.

One approach to monitor deviations from standard medical practice is by assessing adherence to CPGs. CPGs are collective sets of treatment recommendations that attempt to capture the best medical practices for different pathologies [5]. CPGs are promoted as a means to decrease inappropriate practice variation and reduce medical errors [6]. It is generally thought that clinician adherence to CPG recommendations is the primary means to achieve this goal. High levels of adherence to CPGs may indicate optimal care, whereas low adherence rates may suggest sub-optimal treatment. In reality, however, deviation from CPGs often reflects the fact that CPGs cannot be exhaustive; it is not feasible to cover the entire combinatorial space of patient parameters. Deviations from CPG recommendations may thus be beneficial, and it is expected that clinicians will use their personal judgement to contextualize individual patient decisions. In light of the above, previous work identified several barriers to adherence including physician familiarity with the CPGs, physician attitudes towards the CPGs, environmental factors, CPG implementation factors and patient-related factors such as preference [7,8].

Monitoring compliance to CPGs in the clinical setting can be labor intensive. Therefore, in this study we strived to automate the characterization of adherence to CPGs using natural language processing, data modeling and comparison algorithms. Our vision was to computationally parse electronic health records (EHRs) containing both structured and unstructured data to quantify adherence levels, categorize the types of deviations from CPG recommendations, and finally identify the potential rationale for these deviations.

We demonstrate our approach using EHRs of patients diagnosed with adult soft-tissue sarcoma (STS). STS is a group of connective-tissue based cancers that account for roughly 1% of new cancer diagnoses with historical five year survival rates of slightly greater than 50% [9]. These cancers have diverse anatomical origins and can derive from multiple somatic cell types. The variety of histologies results in the presence of multiple drug options and the different anatomical locations offer multiple surgical possibilities. As STSs are rare cancers with numerous treatment options, it is not surprising that prescriptions for patients frequently deviate from CPGs, making STS an ideal use case to evaluate our methodology [10,11,12].

Methods

Description of concepts

The CPGs used in this study were developed by the Lombardy Oncology Network, a data sharing network that contains over

fifty care premises in Northern Italy. Patient data used in this work were gathered at the Fondazione IRCCS Istituto Nazionale dei Tumori (INT), a network member and thought leader, from November 2006 to November 2012.

The CPGs contained hundreds of clinical cancer presentations (conceptually similar to diagnosis) with matching recommended treatments. There are multiple recommended treatments for each clinical presentation. A single CPG recommendation was defined as the unique coupling of clinical presentation, recommended treatment, and start/end date. The study involved 1484 separate CPG recommendations.

Individual clinical presentations were modeled as a data structure of the following clinical fields: tumor anatomic location, tumor depth (deep/superficial), tumor grade, tumor size, disease status, tumor histological type (liposarcoma, etc.) and surgical status (tumor resectable/not resectable). A clinical presentation included all or a subset of the fields. This modeling approach is standard for CPGs, and is similar to that used by the National Comprehensive Cancer Network [13]. An example of an STS clinical presentation in the Lombardy CPGs is: "Patient with adult soft tissue sarcoma located in the limb or torso with a deep, high grade, ≥5cm, localized tumor".

Treatment programs (TPs) were defined as sequences of medical procedures (treatment elements), for example "Wide surgical excision with adjuvant/neo-adjuvant radiotherapy". A treatment element contained items such as drug administration, surgery, radiotherapy, and transplantation. The Lombardy CPGs contained recommended TPs for each clinical presentation.

Once a physician selected a particular clinical presentation from the CPGs, the matching TPs were presented via the local EHR system. Physicians were able to prescribe a TP that was discordant with CPG recommendations (Figure 1). In doing so, they entered their alternative TP in free-text form. The EHR system recorded this decision as well as additional relevant notes provided by the physicians.

Data regarding the treatment was also entered into the EHR system by caregivers during TP execution. We applied standard NLP methods on this data to deduce the actual TP that a patient underwent. The extracted TP was compared to the CPG recommended TPs to assess adherence. The actual TP was considered to deviate if it was discordant to CPG recommendations, regardless of whether the prescription was according to the recommended TPs or not (Figure 1).

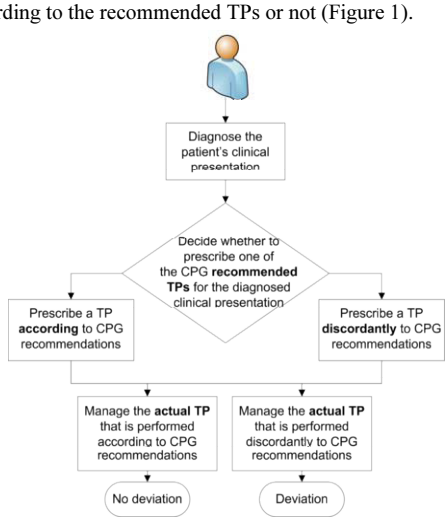


Figure 1 - CPG assisted decision making.

Application of NLP techniques

We applied NLP techniques on the EHR free text data to computationally retrieve the required information for this study. After Italian to English machine translation, we used the Unstructured Information Management Architecture (UIMA) framework to process unstructured information [14]. Our UIMA pipeline included tokenization, parts of speech (POS) tagging, normalization using standard terminologies in UMLS [15], entity and relationship extraction, semantic analysis, negation and disambiguation reasoning. Resulting structured annotations included drugs, diseases, procedures, symptoms, body regions and tumor characteristics. Relationship extractions were used to infer aspects such as the number of chemotherapy cycles, tumor size, tumor grade, and reasons for specific treatment prescription.

The IBM Advanced Care Insights platform (ACI) was used to run the UIMA framework. Within ACI, IBM Content Analytics Studio (ICA Studio) was used to build a Processing Engine Archive file (PEAR), which is a standard UIMA packaging format that can be deployed within any UIMA-compatible framework.

The basic building blocks of ICA Studio are dictionaries and rules. There are three types of rules: break rules that are used for tokenization, character rules that are used for pattern recognition of specific types of information such as dates, names and units, and parsing rules that are used to analyze tokens, other UIMA annotations, and their relationships.

ACI's built-in medical dictionaries contain RxNorm, SNOMED CT, ICD-9, ICD-10, LOINC, and HL7. We supplemented ACI with dictionaries containing chemotherapy drugs, local clinical studies, tumor parameters, and treatment reasons. Finally, ACI also contains entitiy mappings such as SNOMED CT to ICD-9 and ICD-9 to ICD-10.

Treatment program comparison

The results of the text analytics are structured annotations on the text. The annotations first need to be transformed to a pre-defined data model to enable advanced analyses. We therefore designed an actual TP model that defines the treatment which was given to patients (not shown here). The model was designed to enable comparison with the recommended TPs.

To categorize deviations, we identified the most similar recommended TP in the CPGs. The most similar recommended TP was found by assessing the degree of similarity between recommended and actual TPs. The differences between an actual deviating TP and its most similar recommended TP were classified into different categories.

The comparison approach used in this study had three stages. First, we analyzed the abundance of treatment elements to find extra or missing elements. We next used permutation comparison to detected changes in treatment element sequence. The final step was a comparison of the content of every treatment element itself; the specific properties of treatment element were compared. The third step enabled the detection of different chemotherapy drugs, different numbers of chemotherapy cycles, and different surgery types.

Extracting reasons for deviation

We used the same NLP techniques described above to extract reasons for deviation from CPGs. This was done by identifying relationships between extracted annotations using semantic parsing rules. For example, one can consider the following machine-translated sentence: "In light of extension of illness, the patient's age and preliminary activity of molecule in this particular histotype, starting chemotherapy

with gemcitabine". By detecting that the conjunction "in light of" connects the two parts of the sentence, we deduced that the first part of the sentence describes reasons for the given treatment, whereas the second part ("starting chemotherapy …") describes the treatment itself.

Manual validation

We performed manual validation of our computational results on a subset of randomly selected TPs (see results). Four human validators were exposed to the entire EHR records and CPGs. Different subsets of the validation dataset were allocated to each reviewer and results were compiled.

Results

Study setting, patient selection, and data cleansing

Our patient data contained adult STS patients treated at the Fondazione IRCCS Istituto Nazionale dei Tumori between November 2006 and November 2012. We acquired 5598 electronic patient discharge letters representing 2699 STS treatment programs on a total of 2151 different patients. 948 TPs with missing data were excluded consisting of: TPs that were follow-ups, where the actual TP was unknown, did not have at least one CPG recommendation due to CPG incompleteness, or were clinical studies not mentioned in the CPGs. This resulted in 1751 TPs consisting of 1431 patients for analysis (Table 1).

Table 1– Summary of TPs included in the study

Feature	Value
Number of TPs	1751
Male	957
Female	794
Duration	Nov 2006 – Nov 2012
Age range	18 – 100 (median 57)
Unique patients	1431

Quantifying factors that impact deviation frequency

Our computational approach identified deviations in 48.9% of the actual TPs, meaning that most given treatments were found to not fully comply with the CPG recommended TPs.

We next assessed non-clinical parameter correlation with deviation frequency. Strikingly, 35% of the TPs prescribed according to CPG recommendations in reality deviated from the CPG recommendations (Figure 2A). TPs that were prescribed discordantly to CPG recommendations did in fact deviate in 80% cases. Gender and age (cutoff set at median age) were not associated with deviation frequency.

Upon analysis of clinical parameters (Figure 2B), we observed that all disease and tumor parameters were associated with deviation frequencies, except for tumor location. This analysis portrays an expected trend in which poorer prognostic status (large, high grade, and deep tumors) is linked to substantially higher deviation levels. Indeed, the highest deviation frequency was found in metastatic disease (78%).

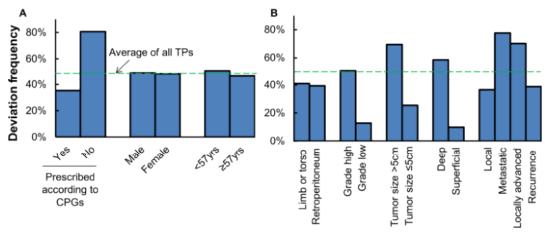


Figure 2 - Features associated with deviation frequency. (A) Demographic/adherence features (B) Clinical parameters. Not all features were specified in each TP. N=: limb/torso (1037), retroperitoneum (304), high gr (661), low gr (512), >5cm (371), ≤5cm (110), deep (445), superficial (331), local (1206), metastatic (359), locally adv. (57), recurrence (183).

Measuring prevalence of different deviation types

TP deviations can be classified using non-mutually exclusive categories. Tabel 2 presents the abundance of deviations that have added or removed treatment elements, exhibited differences in chemotherapy drugs, differences in number of chemotherapy cycles, and differences in surgery type.

The most abundant source of deviation was overtreatment, consisting of 39.7% of all cases in contrast to 12.7% for missing treatments. Notably, metastatic presentations had no excluded elements despite an overall average of 12.7% for all TPs. Also prominent was the observation that there was only a

Table 2 - Frequency of deviation types. Chemotherapy/surgery percentages shown with respect to TPs that included those elements.

Parameters	Added/Removed elements					Chemotherapy differences			Surgery differences	
	N (Dev TPs)	Added element	Missing element	Added and missing	Different order of elements	N (Dev TPs with chemo)	Different drug	Different cycles	N (Dev TPs with surgery)	Different surgery type
Prescribed according to CPGs	431	25.5%	17.6%	0.5%	13.9%	277	34.7%	36.8%	335	32.8%
Prescribed discordantly	426	54%	7.7%	1.6%	7%	301	44.2%	19.6%	239	16.3%
High grade	336	12.8%	24.1%	1.5%	17.6%	220	32.3%	35.5%	278	40.6%
Low grade	66	34.8%	10.6%	4.5%	7.6%	29	3.4%	0%	51	58.8%
>5cm size	258	13.6%	22.5%	0.8%	14%	138	27.5%	41.3%	225	50.2%
≤5cm size	28	46.4%	7.1%	3.6%	7.1%	19	5.3%	15.8%	23	39.1%
Deep	261	13%	23%	1.1%	13%	141	27%	39%	224	52.7%
Superficial	33	42.4%	18.2%	6.1%	15.2%	21	14.3%	14.3%	23	43.5%
Local	441	17.5%	20.6%	1.8%	17%	283	29%	31.4%	355	41.7%
Metastatic	279	57.3%	0%	0%	3.2%	232	10.3%	7.8%	113	0.9%
Loc. advan.	40	27.5%	37.5%	2.5%	12.5%	40	50%	12.5%	19	0%
Recurrence	143	41.3%	9.8%	2.1%	5.6%	98	48%	15.3%	84	16.7%
All TPs	857	39.7%	12.7%	1.1%	10.5%	578	39.6%	27.9%	574	26%

10.3% deviation rate of type 'different chemotherapy drug' for metastatic cases with administered chemotherapy. In general, disease parameters were more strongly associated with chemotherapy differences than surgical differences, with an exception being local/metastatic clinical presentations.

Identification of potential reasons for deviation

NLP parsing identified 1191 potential reasons for deviation among the 857 TPs that we labeled as deviations (average 1.4 per TP). 67.3% of the deviating TPs had one to four reasons and 29.7% had no identified reasons (Figure 3A).

Potential reasons for deviation were classified into five categories: cancer status, other clinical, current treatment related, previous treatment related, and patient preference related. Reasons for deviation that were based on cancer status represented the majority (59%) of all deviations (Figure 3B).

Deviation reasons were further classified into lower-level categories (Figure 3C). The cancer status category consisted of different tumor and disease progression parameters. Other clinically related reasons included demographics, oncological and non-oncological comorbidities, acute symptoms and overall clinical condition. The previous treatment related reasons include amount of previous treatment, poor previous response, previously severe side effects, and presence of residual margins after surgery. The patient preferences category included patient treatment requests or refusals. Lastly, current treatment related reasons consisted of anticipated treatment efficacy, impact on quality of life, and newly available clinical evidence. Deviations due to environmental constraints including lack of personnel or resources were rare and thus not presented.

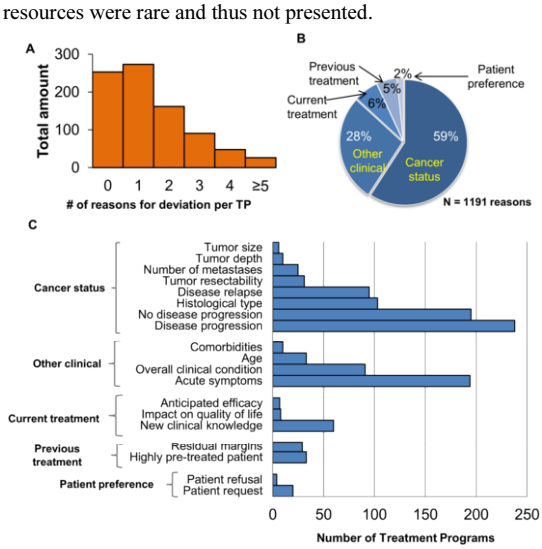


Figure 3 – Potential reasons for deviation. (A) Number of reasons of deviation per TP (N=857 deviating TPs). (B) Different categories of deviations. (C) Deviation subtype analysis (subtypes with n≥4 deviating TPs are shown).

The largest fraction of deviations appear to result from disease progression or a lack thereof, together with presence of acute symptoms. Interestingly, new medical knowledge was only a small fraction of potential deviation causes.

Manual validation

We performed manual validation on a dataset of 222 TPs that were randomly selected from the entire dataset. The validation dataset contained 136 TPs prescribed according to CPG recommendations and 86 discordant TPs.

We first searched for exact matches between NLP-retreived actual TPs to manually retreived TPs. Given this approach, minor or large differences between manually detected and NLP-retrieved actual TPs had the same effect on scoring. The recall for all TPs, TPs prescribed according to CPGs, and discordantly prescribed TPs was 0.71, 0.73, and 0.67, respectively. Chemotherapy cycles were not evaluated for comparison. The TP comparison framework detected differences between actual and CPG recommended TPs. As expected, manual validation showed that the comparison had rather high recall (0.88).

We also assessed deviation labeling. Deviation labeling relies on the accuracy of actual TP extraction and the TP comparison. We compared the quality of our deviation labeling algorithm to a default algorithm that labels TPs prescribed accordingly to CPG recommendations as 'no deviation' and discordantly prescribed TPs as 'deviation' (Table 3). The results show better performance in cases that were prescribed discordantly to the CPGs, while good recall was achieved in the remaining cases.

Finally, a random subset of 60 deviating TPs were selected from the validation dataset for manual analysis of detected reasons for deviation. Validation of detected reasons had the following recall: 0.67, precision: 0.78, and F1 score: 0.72.

Table 3 - Deviation labeling validation (algorithm vs default)

	Entire dataset (n=222)		**TP prescribed according to CPGs (n=136)**		**TP prescribed discordantly to CPGs (n=86)**	
	Alg	Def	Alg	Def	Alg	Def
Recall	0.87	0.62	0.80	NA	0.91	1
Prec.	0.65	0.66	0.47	NA	0.81	0.66
F1 val	0.74	0.64	0.60	NA	0.86	0.79

Discussion

In this work we developed computational techniques to characterize deviations from CPGs in adult STS across thousands of patient records. We identified deviations, classified them by types, and proposed reasons that may reflect the physicians rationale in deviation cases. Beyond the value of understanding clinical deviations, this analysis makes multiple findings that may be useful to sarcoma researchers and the decision support community.

One interesting finding was that approximately half (48.9%) of all TPs deviated from the CPGs. Noting the error present in NLP-based analysis, this value is comparable to a study published in 2012 that reported 54% adherence levels [10] and is higher than a study published eight years prior that had 32% [11]. While being a small sample size, this may suggest that compliance to CPGs is increasing for sarcoma over time.

In contrast to the above, we found that the current deviation level is roughly twice that of a recently published study showing 24% deviation frequency, a study whose data have included many of the same patients as in this work [12]. This discrepancy is due to the fact that deviations in the former study were defined as discordantly prescribed TPs, whereas in this study we analyzed whether the actual given TP deviated. This helps explain the observation that 19% of TPs originally prescribed discordantly to CPG recommendations were not in fact deviations. Conversely, we also found that 35% of all TPs prescribed according to the CPGs were actually deviations.

The observation that adherence to, or deviation from CPGs does not strongly predict if the administered TP adheres to the CPG, may have several causes. For instance, physicians may incautiously select a CPG recommendation with a genuine intent to write the actual prescription in free-text. Alternatively, CPG recommendations may not be clear and their selection may not be sufficiently simple. EHR quality may also play a role if detailed treatment documentation is challenging.

Deviation type analysis identified a large tendency to overtreat patients, representing 39.7% of deviation cases. Overtreatment (reviewed in reference [16]) was 3.1-fold more prevalent than missing treatments. The actual number is likely even higher since we manually observed upon inspection of EHR data that documents are missing for some patients. This issue was addressed by analyzing the treatment program fields, as these tend to summarize the complete course of prior treatment.

Overtreatment was especially evident for metastatic patients (Table 2), whom in general where prone to higher deviation levels (Figure 2). This is interesting when considering that metastatic patients had a relatively small amount of drug related deviations due to the large variety of different chemotherapy options at this disease stage. Deeper investigation of overtreatment cases can potentially assist in reducing treatment cost and improving quality of life.

Mining for deviation reasons can help inform physicians of their medical behavior during the decision making process. This knowledge can motivate specific questions that may impact care, for example why does tumor histological type potentially account for 8.5% of all deviations (Figure 3C).

Several factors that impacted our results warrant mention. First, as with most NLP based studies, our study was also subject to inaccuracies resulting from variability in the free-text representation of medical data. Second, our EHR-based dataset was composed of patient discharge letters, with an average of 2.6 documents per patient. As mentioned, a fraction of patients had missing discharge letters that we were not able to access. We partially addressed this issue by analyzing the treatment program fields. Finally, the study was performed assuming correctness of clinical presentations, although misdiagnosis is expected to be present in oncology care.

Our work focused on developing an integrated framework for understanding physician medical decisions in relation to CPG recommendations. In the future it may be possible to extend this analysis with outcome data, which could subsequently be integrated into EHRs to assist in decision making.

Conclusion

In this study we developed a methodology for a deeper understanding of adherence to CPG recommendations and implemented it in an adult STS use case. The resulting insights from this approach can be used to improve CPGs, to understand the decision making process of physicians, to identify cases where deviations may be beneficial, and to increase adherence to CPGs when deemed appropriate.

Acknowledgments

We thank members of the IBM Biomedical Informatics group and the sarcoma and head & neck departments at the INT for helpful discussions throughout the course of the project.

References

[1] Bodenheimer T. The Movement for Improved Quality in Health Care. N Engl J Med 1999: 340: 488–92.

[2] Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013: 22: ii21–7.

[3] Walsh KE, Dodd KS, Seetharaman K, Roblin DW, Herrinton LJ, Von Worley A, Usmani GN, Baer D, and Gurwitz JH. Medication errors among adults and children with cancer in the outpatient setting. J Clin Oncol Off J Am Soc Clin Oncol 2009: 27: 891–6.

[4] Ziller V, Kalder M, Albert U-S, Holzhauer W, Ziller M, Wagner U, and Hadji P. Adherence to adjuvant endocrine therapy in postmenopausal women with breast cancer. Ann Oncol 2009: 20: 431–6.

[5] Grossman JH and Field MJ. Clinical Practice Guidelines: Directions for a New Program. Nat Academies Pr; 1990.

[6] Lugtenberg M, Burgers JS, and Westert GP. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. Qual Saf Health Care 2009: 18: 385–92.

[7] Francke AL, Smit MC, de Veer AJ, and Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: A systematic meta-review. BMC Med Inform Decis Mak 2008: 8: 38.

[8] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, and Rubin HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999: 282: 1458–65.

[9] Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, and Cronin KA. SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations). Bethesda, MD: National Cancer Institute; 2009.

[10] Perrier L, Buja A, Mastrangelo G, Vecchiato A, Sandonà P, Ducimetière F, Blay JY, Gilly FN, Siani C, Biron P, Ranchère-Vince D, Decouvelaere A, Thiesse P, Bergeron C, Dei Tos AP, Coindre J, Rossi CR, and Ray-Coquard I. Clinicians' adherence versus non adherence to practice guidelines in the management of patients with sarcoma: a cost-effectiveness assessment in two European regions. BMC Health Serv Res 2012: 12: 82.

[11] Ray-Coquard I, Thiesse P, Ranchère-Vince D, Chauvin F, Bobin J-Y, Sunyach MP, Carret JP, Mongodin B, Marec-Bérard P, Philip T, and Blay JY. Conformity to clinical practice guidelines, multidisciplinary management and outcome of treatment for soft tissue sarcomas. Ann Oncol Off J Eur Soc Med Oncol ESMO 2004: 15: 307–15.

[12] Waks Z, Goldbraich E, Farkash A, Torresani M, Bertulli R, Restifo N, Locatelli P, Casali P, and Carmeli B. Analyzing the "CareGap": Assessing Gaps in Adherence to Clinical Guidelines in Adult Soft Tissue Sarcoma. EFMI STC Decis Support. 2013: 186: 46-50.

[13] NCCN Guidelines. http://www.nccn.org/clinical.asp. Natl Compr Cancer Network.

[14] Ferrucci D and Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat Lang Eng 2004: 10: 327–48.

[15] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004: 32: D267–70.

[16] Esserman LJ, Thompson IM, Jr, and Reid B. Overdiagnosis and overtreatment in cancer: An opportunity for improvement. JAMA 2013: 310: 797–8.

Address for correspondence

Dr. Zeev Waks. Email: zeevw@il.ibm.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-285

INITIATE: An Intelligent Adaptive Alert Environment

Borna Jafarpour[a], Samina Raza Abidi[b], Ahmad Marwan Ahmad[a], Syed Sibte Raza Abidi[a]

[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Canada*
[b] *Medical Informatics, Faculty of Medicine, Dalhousie University, Halifax, Canada*

Abstract

Exposure to a large volume of alerts generated by medical Alert Generating Systems (AGS) such as drug-drug interaction softwares or clinical decision support systems overwhelms users and causes alert fatigue in them. Some of alert fatigue effects are ignoring crucial alerts and longer response times. A common approach to avoid alert fatigue is to devise mechanisms in AGS to stop them from generating alerts that are deemed irrelevant. In this paper, we present a novel framework called INITIATE: an INtellIgent adapTIve AlerT Environment to avoid alert fatigue by managing alerts generated by one or more AGS. We have identified and categories the lifecycle of different alerts and have developed alert management logic as per the alerts' lifecycle. Our framework incorporates an ontology that represents the alert management strategy and an alert management engine that executes this strategy. Our alert management framework offers the following features: (1) Adaptability based on users' feedback; (2) Personalization and aggregation of messages; and (3) Connection to Electronic Medical Records by implementing a HL7 Clinical Document Architecture parser.

Keywords:

Alert Fatigue; Alert Management Engine; Alert Management Strategy; Semantic Web; HL7-CDA.

Introduction

In recent years, medical Alert Generating Systems (AGS), such as vital signs monitoring devices, drug-drug interaction systems, and clinical decision support systems, have become prevalent in health-care environments. Extensive exposure to a large volume of alerts, especially irrelevant alerts, overwhelms users of these systems and causes alert fatigue. Alert fatigue leads to ignoring vital alerts, longer response times, anxiety in healthcare professionals and incorrect decisions [1]. Due to these undesirable effects, several attempts have been made to address the issue of alert fatigue. Existing approaches can be classified under two general categories: (a) *Suppressing alerts*: the rules generating alerts in AGS are modified so that irrelevant alerts are generated less frequently. For instance, in drug-drug interaction systems, if route of administration is considered in the head of rule, several alerts pertaining to nonexistent interactions will be avoided. In this approach, since the alert management strategy is implemented in the AGS, it cannot be reused in other AGS. Moreover, this approach is not readily applicable to situations where several AGS are operating concurrently; and (b) *Managing alerts*: in this approach, an alert management strategy is defined for each type of alert. An alert management engine monitors the generated alerts by the AGS and manages their lifecycles based on their management strategy. Hence, alert fatigue is prevented by less frequent and smarter generation of alerts as they can only be raised when they are in critical stages of their lifecycle and certain conditions are met. Since alert management is performed externally, this alert management strategy can scaled across multiple concurrent AGS.

In this paper, we present a novel alert management framework featuring an *Alert Management Strategy Language (AMSL)* represented as a Web Ontology Language (OWL) ontology; and, coupled with an *Alert Management Engine (AME)* that executes the alert management strategy based on patient information accessible from a *Health Level 7* (HL7) compliant Electronic Medical Records (EMR) using the *Clinical Document Architecture* (HL7-CDA [2]). We have identified and categorized the lifecycle of different alerts and have developed alert management logic as per the alerts' lifecycle. Our framework implements a set of unique features to address alerts fatigue, such as a smart alert counter that filters out irrelevant alerts based on their time stamps, alert notification adaptability based on users' feedback, aggregation of alerts to a unified alert, responding to the alert delivery medium and possibility of connecting to commercial EMR systems. We leverage semantic web technologies to represent the alerts and to manage them.

Related Work

Approaches to avoid alert fatigue can be categorized into three categories. In the first approach, the underlying rules generating the alerts are modified to limit the generation of irrelevant/repetitive alerts or to delay the notification of the alerts until a point that the alert becomes critical. For instance, in vital signs monitoring devices, generation of an alert regarding high heart rate can be delayed until a certain amount of time has passed since the heart rate has been more than a predefined threshold [3]. In another example, *Rule1* that represents theophylline–cimetidine interaction in a drug-drug interaction system is modified in terms of *Modified_Rule1* so that route of administration is taken into consideration to make the rule fire less frequently and more accurately [4]:

Rule1: If (theophylline and cimetidine)→ Alert (Name: theophylline –cimetidine, Message: "….")

Modified_Rule1: If (theophylline and (**oral** cimetidine)→ Alert (Name: theophylline –cimetidine, Message: "…")

The approach to modify alert generation rules to address alert fatigue has the following limitations: (1) the alert generation logic is encoded within the rules, and hence, modifications to the rules is local to the AGS and it cannot be reused by other AGS; and (2) if several sources of alerts exist, each AGS rule set needs to be modified which may eventually lead to inconsistencies across the multiple alert generation rules.

The second approach to address alert fatigue is to allow the AGS to generate the alerts, but subsequently filter the irrelevant alerts using an auxiliary alert management engine. To filter irrelevant alerts, domain experts' knowledge regarding relevancy of alerts is captured in terms of a classifier. In this approach, AGS users tag the alerts, as either relevant or irrelevant, in a normal no alert fatigue prevention setting. These tags are, then, used by machine learning algorithms to train a classifier to determine if an alert should be suppressed or presented to the user [5]. Disadvantage of this approach is the need for a large vole of training data that cover the entire breath of the alerts in different clinical scenarios, and the fact that the decision logic of such classifiers cannot be interpreted by healthcare professionals.

The third approach to avoid alert fatigue is to capture domain experts' knowledge in terms of an alert management language and represent the lifecycle of an alert—i.e. describing when and how often an alert should be shown to the user. Klimov et al. [6] have defined a comprehensive language to define alert lifecycles.

In our work, we pursue the third approach as it allows an explicit description of alerts and their lifecycles, which can be used to develop generic AGS. In comparison to Klimov et al. [6], our approach is different in the following ways: we define the concept of a *counter* that filters the alerts based on their currency—i.e. either the alerts are deemed too old to be relevant or are regarded as too soon to the previous alert in order for it to be meaningful. We also enhanced our framework by adding a set of desirable features such as adaptability based on user feedback, aggregation of messages, consistent alert management framework across several AGS operating concurrently, and connecting to EMR to be able to use patient information in alert lifecycle definition.

INITIATE: an INtellIgent adapTIve AlerT Environment

Our alert management framework shown in Figure 1 comprises the following components:

(1) An OWL ontology that represent our Alert Management Strategy Language (AMSL).

(2) Alert Management Engine (AME) that exposes RESTful web services for AGS to submit their generated alerts. This engine manages alerts based on the alert management knowledge encapsulated in an instantiation of the AMSL. To interpret the expressions that make use of lab test result and medication conditions, this engine contains an HL7-CDA parser that evaluates validity of expressions such as *allopurinol on hold* and *Synthroid dosage > 100mg*, which are representable in AMSL.

(3) CouchDB that is a NoSQL (JSON document based) database storing the following items:

a. HL7-CDA documents each representing a patient medical record.

b. Instantiation of the AMSL ontology stored as a RDF/JSON document that represents the lifecycle of alerts generated by the connected AGS.

c. Alert Status: each alert status represents how many times and when a specific alert has been generated for a patient. We will see an example of this later in the paper.

We used a document-based database to avoid normalizing our ontology and HL7-CDA documents into a relational table in order to store the documents as a whole.

Alert Management Strategy Language

In this section, we discuss AMSL, which is formalized using an OWL ontology. Instead of describing the ontology structure, we discuss the grammar of AMSL in the grammar discussed in this section, where [] represents an optional element, * represent an element that can be repeated from *0* to *n* number of times, and terminals are put between double quotes. An alert management strategy is represented by a sextuple:

Alert → ([AGSName], AlertName, [LifeCycle], [AlertConfig], PatientID*, UserID*).

UserID and *PatientID* show the list of users and patients that use this strategy for this alert. *AlertConfig* and *LifeCycle* are discussed in the rest of this section. *AGSName* names the AGS that is the source of the generated alert.

LifeCycle

LifeCycle represents the lifecycle of alerts and is defined as follows:

- LifeCycle → ([CfounterConfig], [SeverityConfig], [ActivationExpression], [InactivationExpression]).

CounterConfig of a *LifeCycle* defines a counter that represents the number of times an alert has been generated for a patient.

- CounterConfig → ([ResetDuration], [DurationBeforeIncrease]).

Alerts that are too old are deemed to be irrelevant, and hence should not be considered in the counter of that alert. The logic is as follows: if the time when the alert was last generated is more than *ResetDuration*, the counter for that alert is set to zero. Likewise, alerts that are generated too close to the previous one may be irrelevant as well. To account for this situation, if the duration between the last time the counter was incremented and the generation of the current alert is less than *DurationBeforeIncrease*, the counter will not be increased. As we will describe later, this counter is used in the calculation of severity and evaluation of activation and inactivation conditions.

- SeverityConfig→ ([MinSeverity], [MaxSeverity], [NumOfSteps]).

Severity of an alert during its lifecycle may change based on its counter number. Severity changes from *MinSeverity* to *MaxSeverity* in *NumOfSteps* steps as the counter increases. For instance, if the severity configuration is (0,1,10), severity of alert will linearly increase from 0 to 1 in 10 steps.

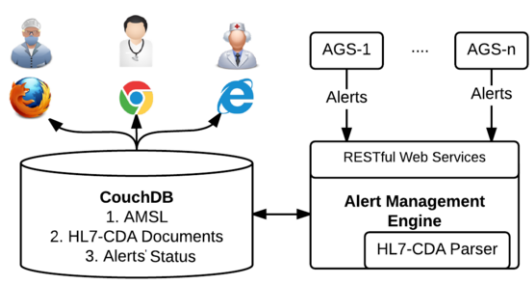


Figure 1 – Our Alert Management Framework

To avoid alert fatigue, an alert generated for a patient can be suppressed until the alert (or the patient) meets certain criteria. Until these criteria have not been met, the alert will not be activated (i.e. notified to the user). As an example, when a patient enters a high value for the blood pressure it will lead to generation of an alert. However, one high blood pressure value does not warrant immediate attention, as it may be a misreading or inaccurate measurement, hence such an alert can be suppressed, and only high blood pressure alerts that are generated within a specific time span may need medical attention. Another mechanism to avoid alert fatigue is to inactivate alerts that are not relevant anymore. An alert may be deemed irrelevant because its purpose has been accomplished. For instance, an active alert "Warfarin is recommended" can automatically become inactivated (i.e. hidden from user) when the patient is prescribed Warfarin. AME uses *ActivationExpression* and *InactivationExpression* in order to activate or inactivate an alert. An expression is defined as follows.

- Expression → Condition | (Expression | Condition, Operator, Expression | Condition).
- Operator → "or" | "and" .

Several types of conditions based on medications of a patient, lab test results, counter of the generated alert, or its severity may be defined:

- Condition → CounterCondition | MedicationCondition | SeverityCondition | LabResultCondition.

A *CounterCondition* is defined as follows:

- CounterCondition → (Comparator, Threshold).
- Comparator → "<" | ">" | "=" | "<=" | ">=".

Activation and inactivation of alerts may be based on their counter. For instance counter condition {"Comparator" : ">" , "Threshold" : 5} will be satisfied for a condition when its counter is 6 or greater.

- SeverityCondition → (Comparator, Threshold).

As we discussed previously, severity of a condition may change as the counter of the alert is changing. For instance, the severity condition (<, 0.5) is satisfied if severity of the generated alert is less than 0.5.

- MedicationCondition → (MedStatus | DosageChange | (Comparator, Threshold), MedicationName | MedicationCategory) .
- MedStatus → "active" | "onHold".

We can define conditions based on medications that exist in the health record of the patient. For instance, we can define a medication condition that is satisfied when a specific medication is *on hold*. Upon satisfaction of this condition based on patient record, the corresponding alert will be inactivated and removed from the list of active alerts shown to the user.

- LabTestCondition → ("existence" | (Comparator, threshold, unit) | TimesUpperLimit), labTestName.

Conditions based on lab test information of patients can also be defined. This type of condition can check existence of a specific lab test and whether its result is less or more than a specific threshold. The *TimesUpperLimit* value shows how many times value of the lab test result should be greater than the upper normal limit so that the corresponding condition is satisfied. For instance, {"existence", "labTestName" : "TSH"} will be satisfied if a TSH test result is present in the electronic medical record of the patient.

AlertConfig

AlertConfig represents how alerts are communicated with the user and aggregated in our framework. Receiving several alerts in one communication can potentially reduce the alert fatigue as relevant alerts are communicated and attended together.

- AlertConfig→ (CommunicationMedium, [AggregationConfig])
- CommunicationMedium → "SMS" | "e-mail" | "web-dashboard".
- AggregationConfig → ("non-aggregatable" | ("aggregatable", maxWait4Aggregation)).

CommunicationMedium represents the medium of communication for the generated alert. Alerts sent through a common communication medium can be aggregated if they are "aggregatable". *MaxWaitForAggregation* shows the maximum amount of time that sending a message can be delayed so that it can be aggregated with other messages. We will see an example of *AlertConfig* and alert aggregation in the next section.

Alert Management Engine (AME)

AME exposes RESTful web services that can be used by AGS to submit their generated alerts. Each incoming alert is a quadruple representing the AGS that has generated the alert, patient ID, the generated alert, and its time stamp. Incoming alerts will be used to update the corresponding JSON documents in CouchDB that represent the history and status of that alert-patient-user combination. In the rest of this section, we describe components and capabilities of AME.

Alert Management Algorithm

Upon receiving and incoming alert by AME, the following steps are performed in order to manage an alert:

1. Update alert's time stamps in CouchDB.
2. Update the alert's counter based on the time stamps.
3. Update alert's severity based on the counter.
4. Evaluate conditions.
5. Evaluate expressions based on conditions.
6. Alter state of alerts based on satisfaction of Inactivation and Activation Expressions.
7. Perform aggregation if possible.

HL7-CDA Parser

Medication and lab test conditions are evaluated based on patients' records. To connect to EMR, we assume patient records are received in Clinical Document Architecture (CDA) documents that is an XML based HL7 standard. HL7-CDA documents represent a snapshot of the patients' record at the time of creation. HL7-CDA documents are transformed to JSON and stored in CouchDB. HL7-CDA parser uses the Java JSON library to evaluate medication and lab test conditions. AME uses these evaluation results to calculate satisfaction of medication and lab test results conditions.

Adaptability and Personalization

We enable personalization in our framework by allowing each user-patient combination to have a separate configuration for each alert. Adaptability enables an alert management engine to tune its parameters in order to fit the needs of a specific user. Existing AGS provide this capability by enabling users to modify parts of the alert generation rules such as the

threshold for the blood pressure alert rules. Tuning AGS rules can be time-consuming and requires extensive knowledge of the AGS domain. We propose a mechanism to provide adaptability based on user feedback. Users of AME can provide two types of feedback: (1) an alert is being activated too frequently for a patient, and (2) an alert is being activated less frequently than necessary. This feedback can be used to tune the alert lifecycle parameters accordingly in the following two ways:

(a) Based on the received feedback, parameters that are responsible for the activation frequency of the alert are multiplied by a pre-defined coefficient. For instance, if the feedback is that an alert is being activated less frequent than necessary, frequency parameters are modified in the following way: *resetDuration* *= 1.05, *DurationBeforeIncrease* *=0.95, *MinSeverity* *= 1/0.95, *steps* = round (1/0.95 * *step*). Hence, our alert management framework can adapt to users' needs based on the received feedback from them.

(b) If the user is knowledgeable about the internal workings of AME, he can choose a specific lifecycle parameter such as *ResetDuration* or *MinSeverity* to be modified based on his feedback. As a result, that specific lifecycle parameter is multiplied by a predefined coefficient. For instance, if the user indicates that an alert is generated more often than needed because of *ResetDuration* parameter, this value is multiplied by 0.8 so that the counter is reset more often leading to less frequent activation of that alert.

Aggregation

To understand how aggregation is performed, we go through a simple example. Suppose that two alerts are of the same *AlertConfig* as listed below:

```
{"AlertConfig": { "Aggregatable" : "yes" ,
   "AggregationMedium" : "e-mail",
   "maxWait4Aggregation" : "15m"}}
```

Imagine the following alerts are generated for two different patients (PatientIDs = 12 and 33) and are supposed to be sent to a specific physician (UserID = 7):

```
{"alert": "Warfarin is recommended",
 "patientID": 33, "UserID": 7, "Status" : "active"
 "DocumentID" : "1eb7ee96b33fa120c89"}
{"alert": "patient at high risk of bleeding",
 "patientID": 12, "UserID": 7, "Status" : "active"
 "DocumentID" : "9eb9ee96b88fa000c57"}
```

These two alerts are aggregated by creating the following document in CouchDB by AME:

```
{"DocumentType" : "aggregatedAlert"
 "AlertDocumentIDs" : ["9eb9ee96b88fa000c57",
                       "1eb7ee96b33fa120c89"]}
```

Alert Lifecycle Categories

In order to define the lifecycle of alerts in the IMPACT-AF project[1], we defined 3 general categories of alert lifecycles, made a template for each category, and fleshed out that template for each alert in that category. In the remainder of this section, we review these categories.

Category1: Cut Off Point Alteration

This category represents the lifecycle of alerts that compare patient data to a pre-defined cut off value. To avoid alert fatigue, these alerts can be ignored until that patient's value passes an alternative cut off value. As an example, according to the domain expert, output of the rule "*IF bilirubin is greater than its upper normal limit THEN patient might have abnormal liver function*" can be ignored until bilirubin is greater than two times the upper normal limit*:*

```
{"ActivationExpression" :
  {"LabTestCondition" :
      {" TimesUpperLimit" : 2,
       " MedicationName" : "bilirubin"}}
```

Bolded values in the above example can be replaced to represent lifecycle of other alerts in this category. In the same fashion, different alternative cut off values can be defined for dosage of medications

Category2: Medication and Lab Test based Lifecycles

By considering the type and dosage of medications taken by patients, several irrelevant alerts can be suppressed. As an example, according to the following rule, a heart rate below 60 bpm is considered abnormal: "*If (heart rate < 60) → patient's heart rate is abnormal*". However, this alert would be of less clinical significant if it is not associated with taking a rate control medication. Filtering low heart rate alerts that are not accompanied by such a medication will potentially avoid alert fatigue without compromising patients' safety or unnecessarily interrupting clinicians' workflow. We define the corresponding lifecycle as follows for this purpose:

```
{"ActivationExpression" :
  {"MedicationConditon" : {"MedStatus"  : "active" ,
       "MedicationCategory" : "rate control"}}}
```

As long as the above expression is not satisfied, alert *patient's heart rate is abnormal* will not be shown to the user. Bolded values in the above example can be replaced to represent lifecycle of other alerts in this category. As another example of this category, to avoid alert fatigue, alert pertaining to the following rule should be inactivated when patient starts taking warfarin: "*If the patient has CHA2DS2-VASc score >= 2 THEN alert physician that Warfarin is recommended*". To accommodate this alert fatigue prevention strategy, we define the lifecycle of the above rule as follows:

```
{ "InactivationExpression" :
  { "MedicationConditon" :
    {"MedStatus"  : "active" ,
       "MedicationName" : "Warfarin"}}}
```

Category3: Counter and Severity based Lifecycles

Some alerts can be suppressed until they are generated a certain number of times over a predefined period. For instance, Canadian guideline for management of Atrial Fibrillation recommend tailoring dose of heart rate control medications to bring resting heart rate of a patient to < 100 bpm. The corresponding rule is implemented as follows in the IMPACT-AF rule engine: "*If (heartrate > 100) → patient's heart rate is uncontrolled.*" Setting 100 as a cut point will probably fire too frequent alerts if patient is right on the edge. When presented this matter to the Cardiologist acting as the principal investigator in the IMPACT-AF project, he suggested ignoring this alert unless the average heart rate is more than 120 bpm over 72 hours with readings recorded at least 12 hours apart. If 24 hours passes and the rule engine does not generates any alerts, we start counting again. To accommodate this alert fatigue prevention logic, we defined the lifecycle of this alert as follows:

```
{"counterConfig" :  {"ResetDuration" : "24h" ,
  "DurationBeforeIncrease" : "12h"},
  "ActivationExpression" :  {"CounterCondition" :
"{"Comparator" : ">", "Threshold" : 5}}}
```

1 http://impact-af.ca/

Evaluation

To evaluate AMSL, five general practitioners were given a 20 minutes presentation on purpose of this language, its role in our framework, and several real world examples. Subsequently, participants were asked to assign an integer score from the range 1 to 5 (score 1 means that physician does not agree with the statement at all, and score 5 means that she completely agrees with the corresponding statement). Table 1 represents the percentage each score is chosen by the domain expert for the corresponding statement.

Table 1- AMSL Evaluation Statements and Percentages Each Score Is Give to each Statement[2]

Statements	Received Score				
	1	2	3	4	5
1.Purpose of AML elements are clear	0%	0%	0%	40%	60%
2. AML is easy to use	0%	0%	20%	60%	20%
3. AML can represent all aspects of alert fatigue prevention strategies	0%	40%	20%	20%	20%
4. All concepts in AML are useful for alert fatigue prevention	0%	20%	20%	40%	20%

Table 1 shows that majority of the participants find AMSL clear, easy to use, useful for and capable of representing alert fatigue prevention strategies. In the second part of the evaluation, we asked a domain expert to design 10 patient scenarios each composed of a sequence of 15 events making changes to health record of a patient (e.g., addition of a blood work or heart rate). Feeding these events to the IMPACT-AF rule engine generated 63 alerts. AME suppressed 34 of these messages. The remaining 29 alerts were aggregated into 14 atomic and composite alerts. We asked 5 general physicians to assign a number from the range 1 to 5 to statements 1, 2 and 3 for each aggregated alert, suppressed alert, and inactivated alert respectively. Table 2 represents the percentage that each score is given to the corresponding statement by domain experts.

Table 2 – AME Evaluation Statements and Percentages Each Score Is Give to each Statement[2]

Statements	Received Score				
	1	2	3	4	5
1. Aggregation is correct	3%	3%	11%	20%	63%
2. Suppression is correct	0%	1%	7%	21%	71%
3. Inactivation is correct	0%	0%	6%	4%	90%

Table 2 shows that most of the aggregations, suppression, and inactivation of messages have been performed successfully by AME according to the participants.

Conclusion

In this paper, we proposed a novel alert management framework that introduces unique features, such as: (1) a counter that filters effects of alerts that are too old or too new to be relevant, (2) alert severities that change by time, (3) possibility of defining conditions related to patients medications and lab test results, (4) adaptability based on users' feedback, (5) centralized management of alerts in case several AGS are generating alerts concurrently, (6) aggregation of alerts (possibly from different AGS) to reduce the number of alerts shown to the user, and (7) capturing communication medium information.

Moreover, our review of the literature shows that most of the alert management strategies are implemented for rule based decision support systems. Since our alert management framework is AGS dependent and does not make any assumptions regarding alert generation mechanism, it can be used to avoid alert fatigue in a variety of clinical decision support systems that use alternative technologies, such as neural networks or Bayesian belief networks.

An interesting future work is to explore the effects of alert management in general, adaptation, and personalization of alerts, in particular on patients' safety. For instance, in cases where physicians are heavily reliant on specific alerts for patients' safety, unpredictability of alerts due to varying parameters may potentially have perilous effects. In our framework, we defined a minimum severity to avoid this scenario. Other potentially dangerous scenarios should be identified and taken into consideration.

Acknowledgement

This research is supported by a research grant from Bayer Healthcare for the IMPACT-AF project

References

[1] McCoy AB, Thomas EJ, Krousel-Wood M, and Sittig DF. Clinical Decision Support Alert Appropriateness: A Review and Proposal for Improvement. Ochsner J . Ochsner Clinic Foundation 2014: 14(2): 195–202.

[2] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron P V, et al. HL7 clinical document architecture, release 2. J Am Med Informatics Assoc 2006; 13(1): 30–9.

[3] Graham KC, and Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010: 19(1): 28–34.

[4] Van der Sijs H, Aarts J, Vulto A, and Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006: 13(2): 138–47.

[5] Imhoff M, Kuhls S, Gather U, and Fried R. Smart alarms from medical devices in the OR and ICU. Best Pract Res Clin Anaesthesiol 2009: 23(1): 39–50.

[6] Klimov D, and Shahar Y. iALARM: An Intelligent Alert Language for Activation, Response, and Monitoring of Medical Alerts. In: Riaño D, Lenz R, Miksch S, Peleg M, Reichert M, ten Teije A, eds. Process Support and Knowledge Representation in Health Care SE - 10. Springer International Publishing, 2013; pp. 128–42.

Address for correspondence

Syed Sibte Raza Abidi, NICHE Research Group, Faculty of Computer Science, Dalhousie University, E-mail: ssrabidi@dal.ca

2 Score 1 means that physician does not agree with the corresponding statement at all and score 5 means that he completely agrees with it

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-290

Influence Diagram As a Support Tool for Clinical Decisions In Cardiopulmonary And Metabolic Rehabilitation

Patrícia da Silva Klahr[a,b,c], Christian Corrêa Coronel[b], Caroline Cabral Robinson[a], João Marcelo Fonseca[a], Cecília Dias Flores[a], Rodrigo Della Méa Plentz[a]

[a] *Universidade Federal de Ciências da Saúde de Porto Alegre, Brazil*
[b] *Instituto de Cardiologia- Fundação Universitária de Cardiologia do Rio Grande do Sul, Brazil*
[c] *Faculdade de Desenvolvimento do Rio Grande do Sul, Brazil*

Abstract

An influence diagram (ID) is a method of graphical representation of uncertain knowledge, which can be employed to support decisions in health care using probabilistic reasoning. We aimed to describe the development of an ID to support the decision-making process in phase II at Cardiopulmonary and Metabolic Rehabilitation Program (CPMR). The development of the ID was carried out through the identification of relevant variables and their possible values, as well as the identification of details of each variable, in order to find a network structure that appropriately connects the nodes that represent the variables, with arcs linking acyclic graphs, and to build the graph using specialized knowledge and the conditional probability table for each node in the graph. In spite of the complexity of the interactions, the model obtained with the ID seems to contribute in the decision-making process in phase II CPMR, providing a second opinion to the health pratictioner and helping in diagnostic, therapeutic and decision-making processes, since it is useful in situations with non-linear modeling or with absent or uncertain information.

Keywords:
Cardiovascular diseases; clinical decision support systems; influence diagram; physical and rehabilitation medicine, probabilistc networks; rehabilitation.

Introduction

The Influence Diagram (ID) appeared for the first time in the United States during the 1980's as a way to represent a decision-making problem [1]. It is even more compact than a decision tree and can make explicit the probabilistic dependencies among variables. An ID is a graphical structure that allows the modeling of uncertain variables and decisions that explicitly reveal probabilistic dependency in a flow of information [2].

There are several benefits in the evaluation of a problem through ID operations such as: the algorithm executing the entire inference and analysis automatically; the analysis being made available in a representation which is natural for decision making; and the use of ID resulting in gains in processing, as it considerably reduces the size of intermediary calculations and the need for greater memory spaces.

Recent work has shown the viability in the use of ID in processes where variables are uncertain and decisions need to be taken starting from the probabilistic dependency in a flow of information as, for instance, in the Medical field [3-5], and Risk Evaluation [6].

The ID is presented in this paper is the selection of the best Cardiopulmonary and Metabolic Rehabilitation (CPMR) option for cardiac patients chosen in a safe and effective way.

Many studies about the enormous prevalence of heart disease suggested some treatment possibilities to minimize the negative effects of these disorders in a patient's quality of life [7-9]. There emerged the possibility of a non-pharmacological treatment: the Cardiopulmonary and Metabolic Rehabilitation (CPMR) [10, 11].

CPMR is the sum of the activities needed to ensure patients with heart disease achieve better physical, mental and social conditions [8-10]. Patients who adhere to CPMR programs experience improved quality of life, hemodynamic stabilization, metabolic changes, and improved vascular and psychological states, which are associated with better control of risk factors and improvement in lifestyle [15-16]. Studies have already demonstrated the cost-effectiveness of CPMR, which became necessary for the rehabilitation of patients affected by these diseases [8-9, 12].

However, there are several uncertainties surrounding the professionals who work in CPMR. Although the beliefs are strengthened and the uncertainties are reduced over years of experience and professional practice, there will always be a degree of uncertainty in each decision. A classic example observed in clinical practice is the prevalent use of the treadmill, with the use of the stationary bicycle restricted to patients who have a physical or mental limitation that prevents treadmill training. However, there are still uncertainties about whether the equipment chosen by the professional is suitable for the patient's specific clinical picture [11, 13].

Thus, based on the technological advances, the possibility of building decision support systems, the high prevalence of heart disease and the growth and recognition of the CPMR, this study aims to build a ID to support clinical decision for phase II CPMR in cardiac patients. This aims to assist non-specialist professionals in choosing the best CPMR option for cardiac patients in a safe and effective way.

Methods

This is a methodological study approved by the Research Committee (ComPesq) of the Federal University of Health Sciences of Porto Alegre (UFCSPA) under number 011/2013.

Participants

The data used in this study were obtained from the medical records of a cohort of cardiac patients in phase II CPMR, assigned by a referral center for cardiopulmonary and metabolic rehabilitation in the state of Rio Grande do Sul, Brazil.

Influence Diagram development

The methodology used to generate the ID began by reading the selected references and isolating clinical information that may influence the diagnosis [13], prognosis and treatment, or be related to the measures taken. The bibliography search and choice of variables was done by a field expert, with the choice of variables and future connections between them based on his knowledge of the field. This study used the variables listed in the scientific literature through consensus, guidelines and norms for CPMR [7, 10, 14, 17, 18].

The structure building can be carried out manually and was designed precisely this way, taking into account the causal relationships between selected variables [13]. The estimating probabilities was obtained from a cohort of data provided by a referral center for CPMR.

However, this does not invalidate the ID, because remains the exact same principle used by the expert to make the decision in scenarios where the quantitative knowledge of the problem is not known or clear [21, 23-25]. At this point, the cohort data provided by the referral center was critical, because it helped to quantify the frequency of clinical outcomes.

Validation

Data for validation [13] of the ID were obtained from the records of another cohort of patients in phase II CPMR, kindly provided by a referral center for CPMR in the State of Rio Grande do Sul called Instituto de Cardiologia do Rio Grande do Sul. Therefore, to assist with the observed frequencies, both situations were considered: the frequency of observations performed on data from the referral center and the decision based on expert's opinion.

Results

The quantitative data used for modeling the shape ID were obtained from a cohort of patients who attended a referral center from April 2012 to April 2013. Characterization of variables in this sample is described in the tables below.

Table I - Characteristics of the sample of 264 patients in phase II CPMR from a referral center

Variables	Mean	±SD
Age (years)	62	±12.35
Weight (Kg)	81	±16.67
Height (m)	1.65	±0.09
Systolic blood pressure (mmHg)	125	±19.35
Diastolic blood pressure (mmHg)	74	±12.11
Heart rate (bpm)	72	±13.18
Partial pressure of O_2 (%)	96	±10.54
Indirect Vo2	15	±5.61
Maximum voluntary ventilation (L)	51	±23.48
Predicted percentage of maximum voluntary ventilation	90	±46.81
Sits and stands	11	±3.51
Maximum inspiratory pressure	74	±33.08
Predicted percentage of maximum inspiratory pressure	79	±37.95
Maximum expiratory pressure	88	±42.91
Predicted percentage of maximum expiratory pressure	91	±49.29
6-minute walk test (m)	406	108.18
Predicted percentage of 6-minute walk test	82	±35.69
Gender-n(%)		
Male	160	(60.6)
Female	104	(39.4)
Ethnic group - n(%)		
Caucasian	223	(84.5)
Black	32	(12.1)
Other	9	(3.4)
Body Mass Index - n (%)		
Underweight	4	(1.5)
Normal weight	48	(18.2)
Overweight	102	(38.8)
Obesity (Grade I)	71	(26.9)
Obesity (Grade II)	27	(10.2)
Obesity (Grade III)	12	(4.5)
Main complaint -n(%)		
Fatigue	79	(29.9)
Dyspnea	26	(9.8)
Fatigue associated to dyspnea	29	(11)
Pain in lower limbs	13	(4.9)
Thoracic pain	13	(4.9)
Medicated for cardiac condition - n(%)	244	(92.4)
Time from event to recovery - (months)(median - Q1-Q3)	1	(1-3)
Main symptoms - n(%)		
Asymptomatic	72	(27.3)
Dyspnea	115	(43.6)
NYHA CFI	38	(14.4)
NYHA CFII	42	(15.9)
NYHA CFIII	31	(11.7)
NYHA CFIV	4	(1.5)
Cough	64	(24.2)
Orthopnea	7	(2.7)
Paroxysmal nocturnal dyspnea	23	(8.7)
Chest pain	56	(21.2)
Typical chest pain	30	(11.4)
Functional Class I	7	(2.7)
Functional Class II	12	(4.5)
Functional Class III	7	(2.7)
Functional Class IV	4	(1.5)
Palpitations	75	(28.4)
Dizziness	97	(36.7)
Syncope	13	(4.9)
Intermittent claudication	26	(9.8)

Legend: *SD – standard deviation; n=absolute frequency, %=relative frequency.*

Table II – Major etiologies diagnosed in the sample of 264 patients in phase II CPMR reference center

Etiologies	Yes n	%
Ischemic	153	(58.0)
Valvular	11	(4.2)
Myocardiopathic	24	(9.0)
Others	76	(28.8)
Total	264	(100.0)

Legend: *n=absolute frequency, %=relative frequency.*

Table III – Major risk factors presented by 264 patients in phase II CPMR reference center

Risk Factor	Yes n	%
Dyslipidemia	150	56.8
Systemic hypertension	210	79.5
Sedentary	186	70.5
Diabetes mellitus	78	29.5
Obesity	73	27.7
Smoker	17	6.4
Former smoker	116	44.0
Active alcoholism	34	12.9
Ex-alcoholic	16	6.1
Positive family history	173	65.5

Legend: *n=absolute frequency, %=relative frequency.*

From the variables defined according to the guidelines and consensus and the data obtained in the cohort of patients, we obtained the ID below:

Figure 1: Influence diagram structured to assist in decision making to perform cardiopulmonary and metabolic rehabilitation.

The DI can represent the mutual and hierarchical relationships among many variables using probabilistic rules and therefore, in many cases, their application is more suitable for prognosis and diagnosis.

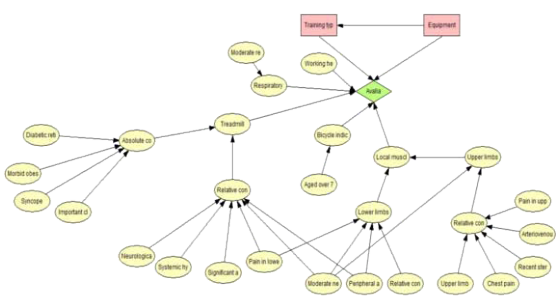


Figure 2: Influence diagram structured to assist in decision making (after entering the cardiopulmonary and metabolic rehabilitation program) regarding type of equipment and training to be used.

We present 3 randomly selected cases from the database to display the variables, the decision of the expert and the decision of the ID.

1. Patient W.L.B., male, 86 years old, 83Kg, 1.60 m height, BMI (body mass index) = 32 kg/m2, resting HR (heart rate) 41bpm. Main complaint is fatigue that comes to feel dizzy, but also refers severe pain in lower and upper limbs. Has positive family history of CAD, is sedentary, stress test is abnormal, predicted maximum inspiratory pressure at 115%. The decision of the expert is to send this patient for the following conduct:. Stretches (flexibility), bike to a HR of 96bpm working with the Borg scale below 3. There is no need for FES (functional electrical stimulation) and IMT (inspiratory muscle training) at the time and would not tolerate the treadmill.
   The decision proposed by the ID was: In failure to perform treadmill, perform bike; indicates the FES, with lower value indicates the RML (located muscle strengthening), does not indicate the IMT and indicates the flexibility. ID also makes reference to the preferable aerobic intensity.
2. Patient L.A.P., male, 48 years old, 76.5 kg, 1.68 m height, BMI (body mass index) = 27kg/m2, resting HR 67 bpm. Main complaint is chest pain. Patient referred not having other symptoms or risk factors such as dyslipidemia, hypertension and family history. Was treated with beta-blockers, anticoagulants and statins; lacked effort and the predicted maximum inspiratory pressure is 105%. The decision of the expert is to send this patient for the following conduct: Stretches (Flexibility); if the patient was not on betablockers, he could be working at a HR of 130bpm, with a range of less than 4 borg There is no need for FES and the IMT. The decision proposed by the ID was: Biggest indication to perform the wake of that bike, indicates the FES, with lower value indicates the RML, and does not indicate the IMT; indicates flexibility. The most indicated type of training is aerobic rather than anaerobic.
3. Patient A. D. S., female, 58 years old, 142Kg, 1.60 m height, BMI = 55Kg/m2, resting HR 136 bpm. AP (arterial pressure) 190/140. Main complaint is fatigue associated with shortness of breath at rest, chest pain, which in some cases reaches fainting and dizziness. Among the risk factors are dyslipidemia, hypertension, decompensated DM, smoking, family history of CAD. Patient states that she has prescriptions but forgets to take the medication. Referred pain in MIE; when the expert did a physical examination he noticed swelling, heat, redness and increased local temperature. The decision of the expert is to send this patient to consult with the specialist because he believes that under such conditions the patient is not able to perform CPMR.
   The decision proposed by ID contraindicated CPMR.

The predictions obtained with this method can inform the patient and the health practitioner whether a given treatment is being effective or whether a given condition is under control.

Discussion

In the simulation of case 1, ID agreed with the expert regarding CPMR and all conducts suggested by the expert are also suggested by the ID, such as opting for bicycle due to advanced age and dizziness presented by the patient. However, the expert, in his daily routine practice, does not

normally use the FES as a therapeutic resource and neglects this feature. Nevertheless, the ID patterned to act as an expert does not rule out the possibility of using the FES, and it indicates the treatment since there is no absolute contraindication [17] for possible use. The ID suggests the most appropriate type of training against HR proposed work, which is too often overlooked to the detriment of a care with the Borg scale and not HR work.

In case 2, the expert, based on information obtained from the patient, opted for the treadmill rather than the bike [19], but the ID, while proposing higher value which realizes the treadmill it, also elected the bicycle as an option. This is because for the ID there is a contraindication to prevent the person from doing bicycle, and only values the use of the treadmill by the patient's condition as favorable for further improvement when using the treadmill. The specialist professional unconsciously and routinely chooses the treadmill over the bike, forgetting that this is also a treatment option in some moments [20]. Since the patient has no inspiratory muscle weakness, both the expert and the ID do not suggest the use of IMT [18, 19].

In the third case, the instability presented by the patient is so great that the expert believed that the risks outweighed the benefits and chose not to indicate the CPMR. The expert guided the patient to return to the doctor and redo the query to stabilize her clinical condition, and then return to CPMR only after that [20,21]. The proposed ID according to the consensus of cardiac rehabilitation when informed with the information available on the patient also indicated that the ideal choice is to not perform CPMR at the moment [8, 10, 11, 14, 16, 22].

In the present study, we attempted to integrate the information obtained in the consensus guidelines, the current scientific literature, and the data obtained in a cohort of patients with the interpretation by experts in order to propose a model to implement a ID. This method explains all the relationships between the predictors and outcomes in a graphical model that incorporates uncertainty through the conditional probability associated with each node [21, 23-25].

The ID subsequently provides the possibility to interpret the relations between the variables and the possibility of intervening in the one that is actually negatively influencing the others. Despite the complexity of interactions, the model obtained in the implementation of an ID seems to be able to adequately describe the relations among the variables.

We understand that the BN and ID can be used in decision support systems in the CPMR. The influence diagram, which is a BN modified [22] for decision making, can provide the necessary tools to generate ideas about the decisions to be made [24, 25].

After defining the outcomes to be considered and the observed frequencies, the ID was modeled to support decision making regarding two moments in the CPMR process. At first, the ID was modeled to help decide whether the patient should perform CPMR. In a second stage, another ID was modeled so that if the answer to the first is "YES", the patient performs CPMR, while the second ID should help to define the modalities of CPMR and what type of equipment and training are best recommended to the studied patients.

Our study was limited regarding validation and statistical analysis of the first ID, since it considers the variables from medical records prior to CPMR and this information was not available in the databank. The databank was exclusive for patients already assigned to CPMR, and theoretically the first ID answered “yes” to follow to the second propagation.

It is expected to continue in the evolution of this research and to associate the ID built to the Simulator of Clinical Cases in Health, in order to obtain a tool to be used in cardiac rehabilitation classes. We also intend to develop an application software for smartphones based on this ID, and make it available to academic students of physical therapy to assist in the learning of phase II CPMR.

Furthermore, our research group aims to improve the development of this ID for CPMR and to develop an ID that addresses all phases of CPMR.

Our group will continue with this form of work and wants to further address the temporal matter. The temporality of the relations between the nodes of the BN is known and generates problems in building them due to false causal effects.

This difficulty is already known and described, and is especially important for networks that express this type of causal relationship between variables (nodes) over time. However, causality can arise from multiple contexts where every node has an influence on a child node. The structural contingencies that may modify these influences, if modified over time, make it more difficult to adequately represent the knowledge in the form of a BN. Many clinical decisions have temporal relationship. Appropriate measures in the initial phase of a condition may be inadequate in the late phase, and vice versa.

Conclusion

It is believed that the ID can significantly contribute to the construction of knowledge, assist in the decision making process and encourage future physical therapists to associate health with biomedical informatics.

Despite the complexity of the interactions, the model for the implementation of the ID seems to be able to predict the scenarios in which the new variables can be incorporated or analyzed, contributes in the health customization process, and ultimately provides a second opinion for the health professional helping the diagnostic, therapeutic process and decision making of the physiotherapist.

Acknowledgments

We thank CAPES for financial support in the form of a scholarship to researcher Patricia da Silva Klahr.

References

[1] Clemen RT. Making hard decisions: An introduction to decision analysis. Duxbury Press, Belmont, CA (1991)

[2] Shachter RD. Evaluating Influence Diagrams. Operat. Res. 34, 871-882 (1986)

[3] Kendrick DC, Bu D, Pan E, Middleton B. Crossing the evidence chasm: building evidence bridges from process changes to clinical outcomes. J. Am Med. Inform. Assoc. 14, 329-339 (2007)

[4] Gómes M, Bielza C, Pozo JAF, Ríos-Insua S. A Graphical decision – theoretic model for neonatal Jaundice. Med. Decis. Making 27, 250-265 (2007)

[5] Lee RC, Ekaette E, Kelly KL, Craighead P, Newcomb C, Dunscombe P. Implications of cancer staging uncertainties in radiation therapy decisions. Med. Decis. Making 26, 226-238 (2006)

[6] Liu H, Zhang C, Wan Y. Research on information engineering surveillance risk evaluation based on probabilistic influence diagram. In: 2nd IEEE International Conference on Information Management and Engineering, ICIME 2010. vol. 5, pp. 362-366(2010).

[7] Fletcher GF, Ades PA, Kligfield P. Exercise Standards dor Testing and Training: A scientific Statement from the American Heart Association. Circulation Journal of the Americam Heart Association (2013) 872-934.

[8] Taylor RS, Brown A, Ebrahin S, Jolliffe J, Nooroni H, Rees K, et al. Exercise-based cardiac rehabilitation in patients with coronary heart disease: systematic review and meta-analysisof randomized controlled trials. The american journal of medicine. 2004, 116 (1): 682-92.

[9] Cortes-Bergoderi M, Lopez-Jimenez F, Herdy AH, Zeballos C, Anchique C, Santibañez C, et al. Availability and characteristics of cardiovascular rehabilitation programs in South America. J. Cardiopulm Rehabil Prev. 2013 Jan-Feb;33(1):33-41.

[10] Carvalho T, Cortez AA, Ferraz A, Nóbrega ACL, Brunetto AF, Herdy AH, et al. Diretriz de reabilitação cardiopulmonar e metabólica: aspectos práticos e responsabilidades. Arq Bras Cardiol. 2006; 86(1):74-82.

[11] Perk J, Backer GD, Gohlke H, Graham I, Reiner Z, Verschuren WMM, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). European Heart Journal. 2012; 33(1): 1635–1701.

[12] Wong WP, Feng J, Pwee KH, Lim J. A systematic review of economic evaluations of cardiac rehabilitation. BMC Health Services Research 2012;12:243.

[13] Flores CD, Höher CL, Ladeira M, Vicari RM. Uma experiência do uso de redes probabilísticas no diagnóstico médico. Argentine Symposium on Artificial Intelligence, 2000.

[14] Ghorayeb N, Costa RVC, Castro I, Daher DJ, Oliveira Filho JÁ, Oliveira MAB. et al. Diretriz em Cardiologia do Esporte e do Exercício da Sociedade Brasileira de Cardiologia e da Sociedade Brasileira de Medicina do Esporte. Arq Bras Cardiol 2013, 100(1Supl.2):1-41.

[15] Moreira PBB. Custo-efetividade de programas de reabilitação cardiovascular. Jornal do Departamento de Ergometria e Reabilitação Cardíaca. 2004; 27: 14-6.

[16] Muela HCS, Bassan R, Serra SM. Avaliação dos Benefícios Funcionais de um Programa de Reabilitação Cardíaca. Rev Bras Cardiol. 2011;24(4):241-250 julho/agosto.

[17] Smart NA, Dieberg G, Giallaura F. Functional electrical stimulation for chronic heart failure: A meta-analysis. International. Journal of Cardiology 167 (2013) 80–86.

[18]Smart NA, Giallauria F, Dieberg G. Efficacy of inspiratory muscle training in chronic heart failure patients: A systematic review and meta-analysis. Int J Cardiol (2012).

[19] Marco E, Ramirez-Sarmiento AL, Coloma A, Sartor M, Comin-Colet J, Vila J, et al. High-intensity vs. sham inspiratory muscle training in patients with chronic heart failure: a prospective randomized trial. European Journal of Heart Failure Advance Access published March 19, 2013.

[20] Araújo CGS, Carvalho T, Castro CLB, Costa RV, Moraes RS, Oliveira Filho JA. Normatização dos Equipamentos e Técnicas da Reabilitação Cardiovascular Supervisionada. Arq Bras Cardiol. 2004;83(5):448-52.

[21] Berchialla P, Foltran F, Bigi R, Gregori D. Integrating stress-related ventricular funtional and angiographic data in preventive cardiology: a unified approach implementing a Bayesian Network. Journal of Evaluation in Clinical Practice 18 (2012) 637-643.

[22] Moraes RS, Nóbrega ACL, Castro RRT, Negrão CE, Stein R, Serra SM, et al. Diretriz de Reabilitação Cardíaca. Arq Bras Cardiol. 2005; 84(5):431-440.

[23] Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013:346:f657.

[24] Verduijn M, Peek N, Rosseel PMJ, Jonge E, Mol BAJM. Prognostic Bayesian networks I: Rationale, learning procedure, and clinical use. Journal of Biomedical Informatics 40 (2007) 609-618.

[25] Vicari RM, Flores CD, Silvestre AM, Seixas LJ, Ladeira; M, Coelho H. A multi-agent inteligente environment for medical knowlwdge. Artificial Intelligence in Medicine 27 (2003) 335-366.

Address for correspondence

Rodrigo Della Méa Plentz. Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA). Rua Sarmento Leite, 245, Porto Alegre, Rio Grande do Sul, Brazil. Post Code 90050-170. Phone +55 (51) 3303-9000. Fax +55 (51) 3303-8810. e-mail: rodrigop@ufcspa.edu.br.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-295

Advances In Infection Surveillance and Clinical Decision Support With Fuzzy Sets and Fuzzy Logic

Walter Koller[a], Jeroen S. de Bruin[b], Andrea Rappelsberger[b], Klaus-Peter Adlassnig[b,c]

[a] *Clinical Institute of Hospital Hygiene, Medical University of Vienna and Vienna General Hospital, Austria*
[b] *Section for Medical Expert and Knowledge-Based Systems, CeMSIIS, Medical University of Vienna, Austria*
[c] *Medexter Healthcare GmbH, Vienna, Austria*

Abstract

By the use of extended intelligent information technology tools for fully automated healthcare-associated infection (HAI) surveillance, clinicians can be informed and alerted about the emergence of infection-related conditions in their patients. Moni—a system for monitoring nosocomial infections in intensive care units for adult and neonatal patients—employs knowledge bases that were written with extensive use of fuzzy sets and fuzzy logic, allowing the inherent un-sharpness of clinical terms and the inherent uncertainty of clinical conclusions to be a part of Moni's output. Thus, linguistic as well as propositional uncertainty became a part of Moni, which can now report retrospectively on HAIs according to traditional crisp HAI surveillance definitions, as well as support clinical bedside work by more complex crisp and fuzzy alerts and reminders. This improved approach can bridge the gap between classical retrospective surveillance of HAIs and ongoing prospective clinical-decision-oriented HAI support.

Keywords:
Healthcare-associated infections; Automated surveillance and monitoring; Clinical alerts; Fuzzy sets and fuzzy logic; Moni.

Introduction

During the last two decades, healthcare-associated infection (HAI) surveillance evolved from rather limited research projects to widespread mandatory quality management (QM). Since HAI surveillance depends on clinical data collected by experienced healthcare workers and infection control practitioners (ICPs), it requires human resources that are usually not continuously available or interfere with other equally important healthcare requirements.

Provided that relevant data can be automatically collected and analyzed by intelligent software, fewer human resources are required to make surveillance data available for infection prevention and control (IP&C), adverse event monitoring, and other fields of patient safety and clinical QM.

Now that systems for IP&C have been established and become more sophisticated, we seem to have reached a point where the actual benefit of these measures is questionable and needs to be assessed [1]. Although HAI surveillance and control have a strong ethical background, the demand for evaluating IP&C in terms of its cost-effectiveness and its relevance in health economy is justified.

This infers a higher level of complexity, calling for new methodological measures to deal with the system appropriately and economically. Here information technology (IT) will play a significant role. With the transformation and results of the Moni system for IT-backed clinical HAI monitoring, we show the potential of this new approach.

Moni was initially designed as a bacteria and antimicrobial resistance monitoring tool. Later it was expanded to clinical data acquisition and was attached to a complex knowledge-processing rule engine, allowing for the interpretation and aggregation of clinical and laboratory findings in patient records. It was intended for automated detection and surveillance of HAIs and their retrospective reporting [2].

From our clinical colleagues we learned that the generally accepted international HAI definitions are of limited value in routine bedside work. Using almost the same dataset, they prefer sensing changes in their patients' clinical condition in a prospective patient-specific manner. This implied extension of terms and definitions in the existing surveillance systems. In particular, uncertainties in definitions and propositions must be considered when trying to assess borderline cases. To address these aspects, fuzzy-set- and fuzzy-logic-based technology was implemented for data and rule processing in Moni. This fundamentally new approach opened new perspectives.

Methods

Moni IT system

Moni is a system for automated nosocomial infection monitoring and surveys, alerts, and reports. This large-scale IT system is in operation at the Vienna General Hospital (VGH), the main teaching hospital of the Medical University of Vienna, with approximately 2,100 beds and 14 intensive care units (ICUs).

The system provides two separate clinical applications: Moni-ICU for adult patients and Moni-NICU for neonates. Moni's main characteristics are fully automated data transfer from electronic sources, the implementation of medical knowledge bases, and the use of specific processing algorithms that rely on these knowledge bases and evaluate, aggregate, and interpret clinical data in a stepwise manner until patient data can be mapped to the included HAI definitions.

Data sources

Moni's data sources are the ICU's patient data management systems (PDMSs), which include administrative and clinical patient data as well as data from the central medical-chemical laboratory information system (LIS). In addition, microbiology data are derived from a separate LIS. All data are sent through a communication interface to Moni's data warehouse, which is the starting point of all subsequent data processing steps.

HAI surveillance and alert definitions

Moni's surveillance is based on the criteria and textual definitions issued by CDC/NHSN, Atlanta [3], ECDC, Stockholm [4], and the KISS criteria issued by the German National Reference Center for Surveillance of Nosocomial Infections, Berlin [5]. Moni's alert criteria are based on established surveillance rules, but were adapted to clinical needs. For instance, parameter definitions of clinical concepts were changed to avoid over-alerting [6].

Arden Syntax

The HAI surveillance and alert definitions were translated into Arden Syntax, version 2.7 [7], by a small team of clinical informaticians and experienced infection control specialists and clinicians. Arden Syntax is a medical knowledge representation and processing scheme for the development of clinical decision support systems. The syntax is maintained by Health Level Seven International [8], and new versions are regularly approved by the American National Standards Institute. A recent report on the Arden Syntax software (compiler, engine, server, database connector) embedded into state-of-the-art service-oriented software architecture can be found in [9].

Data and knowledge processing

Administrative, clinical, laboratory, and microbiological patient data from the PDMSs and the LISs are processed in a stepwise pipeline of aggregation, interpretation, and evaluation [10]. The first steps are pre-processing (checking for missing data and plausibility) and feature extraction (e.g., calculation of means and scores). This is followed by interpretation and intermediate to high-level clinical concept evaluation, for which a package of hierarchically interwoven medical logic modules (MLMs)—the building blocks of Arden-Syntax-based knowledge representation—was established. The topmost clinical concepts are the HAI surveillance and alert definitions to be evaluated. Many of the encoded clinical entities are modeled as fuzzy sets and the inference steps are performed by applying fuzzy logic.

The result of this automated processing is a list of HAIs whose definitions are either fulfilled or fulfilled to a certain degree. This evaluation is based on all available patient data, not only of the selected day but dating back to seven days. HAIs whose definitions are not fulfilled are not displayed.

Fuzzy terms and fuzzy logic

When trying to model uncertainty in clinical medicine, a distinction must be made between linguistic and propositional uncertainty.

Most clinical entities, such as fever or leukopenia, are linguistic concepts with gradual or fuzzy transitions between normal and pathological states—this is known as linguistic uncertainty. Such entities are formally modeled by fuzzy sets introducing a borderline range or transition zone between neighboring concepts.

Propositional uncertainty, in contrast, refers to uncertainty in medical conclusions. This is modeled by truth values between zero and one. One example is a patient in intensive care whose body temperature is in the normal range, but the patient is on external thermoregulation. He/She needs external cooling to decrease his/her body temperature and reduce risk. Since we have a strong but indirect hint towards "fever", the proposition "if thermoregulation then fever" is assigned a truth value smaller than 1.0 (here we assigned 0.8).

Table 1 shows the thresholds of the clinical concepts *increased body temperature, increased C-reactive protein, leukopenia,* and *leukocytosis*. The column "pathological range" contains thresholds as defined in the HAI surveillance definitions published by CDC/NHSN [3], ECDC [4], and KISS [5]. Data assigned to this column are constitutive elements of the surveillance results of these HAIs—provided the other relevant clinical concepts for this infection also fall into this definite pathological range.

The column "borderline range" is an additional range for borderline values and for assigning fuzzy degrees to the respective concept under consideration. By doing so, a body temperature of 37.9 °C signifies increased body temperature to the degree of 0.8. This column will include those patient days for which the clinical concept evaluation yields a near-pathological value, but misses the crisp thresholds defined for classical surveillance.

In classical surveillance, all data in the "normal range" and under the "borderline range" will be regarded as non-pathological and not concurring with the respective HAI surveillance definition. They are cut by a crisp threshold into "black and white". In contrast, the HAI alert criteria are furnished with borderline ranges to recognize tendencies and sense clinically important developments in the patient's state.

Table 1 – Four clinical concepts and their definition by fuzzy sets in the Moni-ICU knowledge base

Clinical Concept (Unit)	**Fuzzy Set**		
	Normal Range	**Borderline Range**	**Pathological Range**
Increased body temperature (fever) (°C)	< 37.5	37.5 – 38.0 [1)]	> 38.0 [2)]
Increased C-reactive protein (CRP) (mg/dl)	< 1.0	1.0 – 6.0 [3)]	> 6.0 [3)]
Leukopenia (WBC/mm^3)	> 5,000	4,000 – 5,000 [4)]	< 4,000 [2)]
Leukocytosis (WBC/mm^3)	< 11,000	11,000 – 12,000 [4)]	> 12,000 [2)]

[1)] as defined by clinicians
[2)] as defined by CDC/NHSN [3], ECDC [4], and KISS [5] for retrospective surveillance purposes
[3)] as defined by clinicians; CRP is an early phase protein, useful as an "infection radar" for prospective purposes
[4)] as defined by clinicians; white blood cell count (WBC) is a slowly reacting indicator, important for surveillance purposes

Study on fuzzy terms and fuzzy logic

In a retrospective study encompassing data for the whole of the year 2011, we analyzed all patient stays in ten ICUs for adult patients of the VGH. All patients aged 18 years or older were included, except admissions for less than 48 hours and the first 48 hours of longer admissions [4]. All clinical concepts and all included HAI surveillance definitions were established according to [4]. However, the borderline ranges were defined by clinical experts from the VGH in our group, referred to as clinicians.

The aim of the study was to determine the extent of borderline pathological states and borderline HAI results in the population under investigation, as a result of applying fuzzy terms and fuzzy logic.

For the analysis, data filtering and cleaning were done in both Python and Microsoft Excel 2007.

Results

Moni's routine operation

At present Moni is established at the VGH for ten ICUs with 87 beds for adults, and four neonatal ICUs with 51 positions. The ICUs and the microbiology department provide about 30,000 relevant data items for Moni-ICU and Moni-NICU every day. For each of the 138 patients, the available knowledge packages containing 72 MLMs for Moni-ICU and 160 MLMs for Moni-NICU are invoked automatically. Clinical concepts, higher-level intermediate clinical concepts, and the topmost HAI definitions are evaluated.

Results of fuzzy terms and fuzzy logic study

During the study period, 2,105 patients stayed at one or more of the selected ICUs; 1,253 patients were male (59.5%) and 844 female (40.1%); gender was not recorded for eight patients. The patients' ages ranged from 18 to 98 years (median 61 years; inter-quartile range 22). In total, 2,429 patient stays were recorded, amounting to 24,487 patient days. The duration of the stay ranged from 2 to 135 days (median 6 days; inter-quartile range 8). Of the 24,487 patient days, 162 were unaccounted for due to export and registration problems, which left 24,325 for data analysis.

For the analysis, we considered the definitions of bloodstream infection (BSI), general central-venous-catheter-associated infection (CRI2), microbiologically confirmed symptomatic urinary tract infection (UTI-A), and not microbiologically confirmed symptomatic urinary tract infection (UTI-B) [4]. Pneumonia was omitted because radiology findings are not yet part of the automated data importation into Moni-ICU. We also did not include local central-venous-catheter-associated infection (CRI1) because its definition in the knowledge base did not contain any fuzzy sets or fuzzy rules.

Table 2 (the first four lines) shows that a significant number of patient data fall into the established borderline ranges (11,474 of 97,300 = 11.8%). This is a relatively high number and indicates a possible near miss in case of HAI surveillance, but a clinically useful indication for HAI alerting.

As shown in Table 2 (last line), in 12.5% of the patient days, at least one of the definitions of the included HAIs (BSI or CRI2 or UTI-A or UTI-B) was fulfilled; a further 2.5% yielded a borderline result, i.e., fulfilled to a certain degree. Notably, in the case of several fulfilled HAI definitions only one was counted, where "present" superseded "borderline" and "borderline" superseded "absent". Moreover, in 606 patient days these data constellations obviously yielded an HAI borderline result. These borderline HAI results are shown, the data are explained fully, and clinicians have the option to explore and judge these results.

Furthermore, the majority of HAI surveillance definitions require microbiology test results. If not available (no specimen for culture taken), classical surveillance would record the respective patient day with "no HAI" despite the presence of clinical signs of infection. Applying "weaker" definitions of HAI surveillance definitions, many of those patient days will fall into the "borderline" category.

Validation of fuzzy sets

Whether the definitions of the fuzzy sets for clinical concepts were adequately chosen to capture those borderline values that rightly generate borderline HAI surveillance results were tested and reported on in a previous study [11]. This investigation was based on central-venous-catheter-associated infections (CRIs) and confirmed, in statistical terms, that most of the expert-defined fuzzy sets and logical constructs currently included in the knowledge base of Moni-ICU are an accurate representation of the CRI borderline patient population.

Clinical studies

Studies addressing the aspect of efficacy have shown the excellent conformance of Moni-ICU with an established clinical reference standard, as well as Moni-ICU's superiority in minimizing time demands on the infection control team with respect to electronically supported versus purely human surveillance. The results can be found in [12, 13].

Table 2 – Frequency distribution of the results of four clinical concepts modeled by fuzzy sets, as well as the topmost HAI definitions (based on a total of 24,325 patient days)

Clinical Concept	Absent n (%)	Borderline n (%)	Present n (%)
Increased body temperature (fever)	16,074 (66.1)	3,421 (14.0)	4,830 (19.9)
Increased C-reactive protein (CRP)	4,383 (18.0)	5,841 (24.0)	14,101 (58.0)
Leukopenia	22,991 (94.5)	668 (2.8)	666 (2.7)
Leukocytosis	15,169 (62.4)	1,544 (6.3)	7,612 (31.3)
BSI or[1)] CRI2 or UTI-A or UTI-B	20,687 (85.0)	606 (2.5)	3,032 (12.5)

[1)] inclusive disjunction with precedence of "present" over "borderline" over "absent"

Table 3 – 3x3 contingency tables for four clinical concepts and the topmost HAI definitions

Clinical Concept		**BSI or[1)] CRI2 or UTI-A or UTI-B**		
		Absent	**Borderline**	**Present**
Increased body temperature (fever)	*Absent*	13,911 (57.2)	484 (2.0)	1,679 (6.9)
	Borderline	2,855 (11.7)	122 (0.5)	444 (1.8)
	Present	3,921 (16.1)	0	909 (3.7)
Increased C-reactive protein (CRP)	*Absent*	4,130 (17.0)	152 (0.6)	101 (0.4)
	Borderline	4,890 (20.1)	453 (1.9)	498 (2.0)
	Present	11,667 (48.0)	1 (0.0)	2,433 (10.0)
Leukopenia	*Absent*	19,578 (80.5)	564 (2.3)	2,849 (11.7)
	Borderline	569 (2.3)	42 (0.2)	57 (0.2)
	Present	540 (2.2)	0	126 (0.5)
Leukocytosis	*Absent*	13,201 (54.3)	551 (2.3)	1,417 (5.8)
	Borderline	1,306 (5.4)	54 (0.2)	184 (0.8)
	Present	6,180 (25.4)	1 (0.0)	1,431 (5.9)

[1)] inclusive disjunction

Discussion

"Crisp" and "fuzzy" definitions

Surveillance as well as benchmarking count on unambiguous consensus definitions, which can easily be shared in networks, but are often compromises negotiated in long expert discussions. Clinicians and intensive care specialists dislike stolid definitions; they prefer a more flexible interpretation of patient data with focus on transition from normal to pathological states, and on borderline situations.

The Moni IT system provides retrospective population-specific infection surveillance related to classical infection surveillance, as well as prospective patient-specific high-level data aggregation for clinical support in early detection of infection. Here, fuzzy logic revealed its potential and aroused the interest as well as cooperation of intensive care specialists.

Benefit of borderline

It should be noted that our analysis counted patient days in which HAI criteria were present and not HAI episodes, each of which usually lasts for several days. International networks report their rates in numbers of HAI episodes per, such as 1,000 patient days or device days (which means days under specific exposure).

Patient days with fever, increased CRP, or leukocytosis were counted more often than days with fulfilled HAI definitions (see last column in Table 2). This can be explained by two facts: a) infection is only one of several conditions which trigger such pathologies, b) HAI definitions are composed of a number of pathologies, often including microbiological verification of the infective agent, all of which must be present. Since microbiological verification of the infective agent is a significant part of many HAI definitions, omission of specimens for microbiological investigations leads to underreporting of HAIs and might explain at least part of the borderline group in our analysis.

The borderline HAI group (2.5% of the totally tested 24,325 patient days) is not very large, but adds another fifth to the number of days with HAIs. It is—by the design of surveillance definitions—not relevant for retrospective surveillance. For the clinician, however (especially when he/she checks Moni's results at least once a day), it might serve as a "radar" for early detection of an incipient HAI. This is very valuable prospective information and a priceless benefit for the individual patient.

The absence of elevated body temperature in 6.9% of patient days, although the patients' data were fully compatible with one of the HAIs (Table 3, first line), might be explained by the frequent use of external thermoregulation for patients with fever (for instance with a cooling blanket). Hence, body temperature was considered normal in patients receiving external thermoregulation. The latter is captured by a separate fuzzy rule, which was not investigated in the present study.

Patient days with fever, increased CRP, leukopenia or leukocytosis fell either into the "absent" or the "present" class of the topmost HAI definitions, but never in the respective "borderline" class—except for one patient day with increased CRP and one with leukocytosis (Table 3, middle column). This can be explained by the following: a) a single clinical concept may or may not be part of a specific HAI definition, and b) a HAI definition is fulfilled when only one of a number of clinical concepts applies (i.e., when the rule defines that either A or B or C is present). We regard this as further proof of the previously reported [12] excellent precision of Moni's interpretation of infection-relevant clinical data.

The numbers and percentages in Table 3 in the borderline categories need further analysis, which will be the subject of another study. It may be stated here, however, that the numbers reflect the previously mentioned difference between clinical concepts as part of HAI definitions and the entire HAI definition itself, as well as different "constituents" of different HAI definitions.

Legal considerations

As long as Moni is applied for surveillance purpose—whether for in-hospital, national, or international quality benchmarking—it is not regarded as a medical device in accordance with the European Medical Device Directive 93/42/EEC. Even when the surveillance functionality of Moni is used by ward physicians at the patient's bedside to evaluate the patient's clinical condition (with potential impact on diagnosis and therapy), the intended use of Moni—as defined by the manufacturer—is strictly confined to surveillance.

Moni's alert component differs in this regard. The medical purpose here—as defined by the manufacturer—is to alert the attending physician, identify infection as early as possible, and incorporate the fact into the patient's treatment and care. Thus, the alert component is a medical device.

At present, the application of Moni's alert section is under clinical evaluation. It is being used and tested by several key clinical users who have received appropriate instructions.

Methodological perspectives

Moni-ICU was designed as a purely knowledge-based system with detailed explanatory capabilities. It does not rely on machine learning. The applied clinical knowledge for HAI surveillance is based on international consensus surveillance criteria, which are well known to the medical community. In addition, extended clinical experience is employed to establish useful alert criteria. The explanatory system clearly states what is included in the system and what is not, thus providing full transparency of any inferred clinical result. The approach of stepwise abstraction and aggregation is very similar to human reasoning in clinical medicine. Fuzzy set theory and logic avoid "jumps" in reasoning.

Conclusion

High precision of surveillance results, fully automated acquisition of relevant clinical and laboratory data, and significant time saving for ICPs have been prominent features of earlier versions of this electronic surveillance tool. Now, with broad application of fuzzy set theory and fuzzy logic, the scope of the system has been widened significantly. Comprehension of the inherent uncertainty of linguistic clinical terms and clinical or practice-based propositions, and thus straightforward "intelligent" processing of patients' borderline values, make it a powerful clinical extension of classical infection surveillance.

Acknowledgments

Alexander Blacky, Angelika Berger and Lukas Unterasinger from the Medical University of Vienna; Rudolf Schuster, Georg Schnizer and Reinhard Krepler from the Vienna General Hospital; and Karsten Fehre, Alexander Seitinger, and Stefan Reich from Medexter Healthcare GmbH.

Conflict of interest

Klaus-Peter Adlassnig is also co-founder, CEO, and Scientific Head of Medexter Healthcare GmbH, Vienna, Austria. Medexter Healthcare GmbH has been established to develop and disseminate decision support systems with confirmed applicability in the clinical setting.

References

[1] Harbarth S. Health-economic evaluation and cost-effectiveness of infection control. Keynote at the 9th Healthcare Infection Society International Conference (HIS2014), 16–18 November 2014, Lyon Convention Centre, France.

[2] Adlassnig KP, Blacky A, and Koller W. Artificial-intelligence-based hospital-acquired infection control. In: Bushko R, ed. Strategy for the Future of Health. Stud Health Technol Inform 149, Amsterdam: IOS Press, 2009; pp. 103-10.

[3] Centers for Disease Control and Prevention (CDC), National Healthcare Safety Network (NHSN): NHSN Patient Safety Component Manual, http://www.cdc.gov/nhsn/TOC_PSCManual_Instructions.html, accessed 10 April 2015.

[4] European Centre for Disease Prevention and Control (ECDC), HAI-net; ICU-acquired infections, http://ecdc.europa.eu/en/activities/surveillance/hai/about_hai-net/pages/icu.aspx, accessed 10 April 2015.

[5] German National Reference Center for Surveillance of Nosocomial Infections, ITS-KISS, http://www.nrz-hygiene.de/en/surveillance/hospital-infection-surveillance-system/its-kiss/, accessed 10 April 2015.

[6] Adlassnig KP, Berger A, Koller W, Blacky A, Mandl H, Unterasinger L, and Rappelsberger A. Healthcare-associated infection surveillance and bedside alerts. In: Hörbst A, Hayn D, Schreier G, Ammenwerth E, eds. eHealth2014 – Health Informatics Meets eHealth, Outcomes Research: The Benefit of Health-IT, Stud Health Technol Inform 198, Amsterdam: IOS Press, 2014; pp. 71-8.

[7] Arden Syntax version 2.7, http://www.hl7.org/implement/standards/product_brief.cfm?product_id=2, accessed 10 April 2015.

[8] Health Level Seven International, http://www.hl7.org/, accessed 10 April 2015.

[9] Adlassnig KP, and Fehre K. Service-oriented Fuzzy-Arden-Syntax-based clinical decision support. Indian J Med Inform 2014: 8(2): 75-9.

[10] Adlassnig KP, Blacky A, Mandl H, Rappelsberger A, and Koller W. Fuzziness in healthcare-associated infection monitoring and surveillance. In: Winston A, Melek W, Hall P, McKenzie S, Gibbs M, Adamson G, eds. Proceedings of the 2014 IEEE Conference on Norbert Wiener in the 21st Century – Driving Technology's Future. Piscataway: Institute of Electrical and Electronics Engineers (IEEE), 86.pdf.

[11] De Bruin JS, Blacky A, Koller W, and Adlassnig KP. Validation of fuzzy sets in an automated detection system for intensive-care-unit-acquired central-venous-catheter-related infections. In: Lehmann CU, Ammenwerth E, Nøhr C, eds. Proceedings of the 14th World Congress on Medical and Health Informatics (MEDINFO 2013), Stud Health Technol Inform 192, Amsterdam: IOS Press, 2013; pp. 215-8.

[12] de Bruin JS, Adlassnig KP, Blacky A, Mandl H, Fehre K, and Koller W. Effectiveness of an automated surveillance system for intensive care unit-acquired infections. J Am Med Inform Assoc 2013: 20(2): 369-72.

[13] Blacky A, Mandl H, Adlassnig KP, and Koller W. Fully automated surveillance of healthcare-associated infections with MONI-ICU – A breakthrough in clinical infection surveillance. Appl Clin Inform 2011: 2(3): 365-72.

Address for correspondence

Professor Walter Koller, Martinstrasse 26, A-1180 Vienna, Austria.
walter.koller@meduniwien.ac.at

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-300

A Decision Fusion Framework for Treatment Recommendation Systems

Jing Mei[a], Haifeng Liu[a], Xiang Li[a], Guotong Xie[a], Yiqin Yu[a]

[a] *IBM Research, Beijing, China*

Abstract

Treatment recommendation is a nontrivial task – it requires not only domain knowledge from evidence-based medicine, but also data insights from descriptive, predictive and prescriptive analysis. A single treatment recommendation system is usually trained or modeled with a limited (size or quality) source. This paper proposes a decision fusion framework, combining both knowledge-driven and data-driven decision engines for treatment recommendation. End users (e.g. using the clinician workstation or mobile apps) could have a comprehensive view of various engines' opinions, as well as the final decision after fusion. For implementation, we leverage several well-known fusion algorithms, such as decision templates and meta classifiers (of logistic and SVM, etc.). Using an outcome-driven evaluation metric, we compare the fusion engine with base engines, and our experimental results show that decision fusion is a promising way towards a more valuable treatment recommendation.

Keywords:

Clinical decision support system, Decision fusion, Treatment recommendation.

Introduction

Clinical Decision Support (CDS) provides recommendations from organized medical knowledge and patient information to improve healthcare delivery. A variety of CDS systems have been proposed and implemented. From the perspective of methodology, some CDS systems are knowledge-driven, such as the computerization of clinical practice guidelines, and some are data-driven, such as the discovery of unknown patterns and trends in a large volume of patient data. From the perspective of analytics, some CDS systems are descriptive, such as getting the most frequently used drugs from similar patients, some are predictive, such as the estimation of clinical outcome for taking a certain drug, and some are prescriptive, such as a long-term planning for the optimal intervention at every possible patient state.

Considering that each CDS system has its own strength and weakness, we might not only leverage a single CDS system, but also apply decision fusion technologies to combine multiple CDS systems' results. Actually, in real life, it is natural to consult "several experts" before making a final decision. As it si said, two heads are better than one. The extensive benefits of decision fusion have been shown up in the Netflix Grand Prize in 2009 [1], which was an open competition to predict user ratings for films, and the winner combined the previous three teams' results to achieve a 10.09% improvement. In addition, the Heritage Health Prize in 2012 [2] was an open competition to predict how many days a patient would spend in a hospital in the next year. Again, fusion methods were widely used by both milestone winners and final winners in this competition.

In literature, the research field of "decision fusion" is known under various names, such as multiple classifier systems, mixture of experts and ensemble learning [3]. A general solution of prior arts is the fusion of different (data-driven) learning algorithms with different parameter settings, e.g., trying various features and datasets. However, for clinical decision support, esp. in evidence-based medcine, not only the data-driven learning algorithms are useful, but also the knowledge-driven modeling techniques are greatly helpful. In this paper, we propose a decision fusion framework for a treatment recommendation system, which combines both knowledge-driven and data-driven approaches.

To be convinced of the benefits of decision fusion, we need evaluation metrics, comparing the results from base decision engines with the results from our fusion engine. However, we observe that state-of-the-art evaluation metrics [4], such as precision, recall and RMSE (Root Mean Square Error), are not applicable for evaluating treatment recommendation systems, due to the *partially observed ground truth*. In machine learning, the term "ground truth" refers to the known facts about the training data set in terms of the learning tasks. Taking the Netflix Grand Prize as an example, the ground truth is the actual user ratings for films, while taking the Heritage Health Prize as another example, the ground truth is the actual days a patient would spend in a hospital in the next year. However, as far as for treatment recommendation, what's the ground truth? A naïve answer might be the actual prescription in real data. But, is that right? Suppose that drug A was recommended by a decision engine, and the physician did choose drug A as the prescription, but unfortunately, the patient outcome of using drug A was bad. In this respect, could we mark the recommendation of drug A as correct? Another story is that drug B was recommended by an analysis module, but the physician chose drug C as the prescription, and the patient outcome of using drug C appeared good. Thus, could we mark the recommendation of drug B as incorrect? What if the patient outcome of using drug B becomes better than using drug C? Therefore, we call it the *partially observed ground truth* (i.e., not all decision options are completely observed with outcomes), and in this paper, we propose an outcome-driven measure for evaluating treatment recommendation systems. Here, we remark that the fusion itself has no impact on the "partialness" of the ground truth. Actaully, it is the "partialness" that brings challenges for evalution of treatment recommendation systems, including the base and fusioned ones.

For experiments, we implement a decision fusion framework, which combines three base decision engines. The first engine is a knowledge-driven engine based on clinical practice

guidelines (CPG [5]). The other two are data-driven engines, of which one is a descriptive analytics based on patient similarity (PSA [6]), and the other engine is a predictive analytics based on outcome prediction (PRE). Using different fusion algorithms (such as decision templates and meta classifiers), we compare the evaluation results and conclude that fusion does help to provide treatment recommedations in a more valuable way.

Methods

We first present a decision fusion framework, followed by the introduction of three base decision engines. Next, we describe the fusion engine with a variety of fusion algorithms. Finally, an outcome-driven evaluation metric is defined, which will be used to compare the fusion engine with base decision engines.

Decision fusion framework

As shown in Fig 1, we propose an open framework for decision fusion. A number of decision engines can contribute to the fusion framework. We implement a fusion engine that gets input from base decision engines. On the client side (such as the clinical workstation or mobile apps), we display the treatment recommendations from each of the component, as well as the final decision from the fusion engine. End users have the privilege to see the outcome from all engines.

The fusion engine has two phases. First is the training phase, where all the results from base decision engines will be fed into a fusion engine to learn a fusion model. Second is the testing phase, where an instance that uses the output of base decision engines as features is created, and the trained fusion model predicts the final outcome. Every base decision engine has been either trained well by its own data source (such as PSA and PRE) or modeled well by its own knowledge source (such as CPG). The training and testing phases are meant for the fusion, rather than any base decision engine.

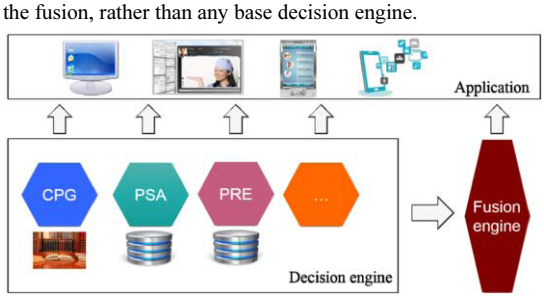


Figure 1 - The decision fusion framework

Decision engine

A decision engine itself provides treatment recommendations, but its methodology and analytical principles may differ from each other. Table 1 shows three different perspectives. The differentiation of decision engines results in the requirement to consult "several engines" before making a final decision.

Technically, CPG [5] is a knowledge-driven decision engine, which computerizes the NICE clinical guideline for Type 2 diabetes. At design time, guidelines are defined as standard (XPDL)-based business processes, where clinical conditions are represented using GELLO expressions. At run-time, a process engine would invoke a query adaptor to retrieve clinical data and a GELLO engine to evaluate clinical conditions whenever a decision-making is needed during the care process. Consequently, clinical recommendations are generated for lifestyle intervention and drug therapy, etc. For example, given a patient who is overweight and whose blood glucose is inadequately controlled by lifestyle interventions alone, CPG would recommend to start Metformin.

Table 1 – Decision engines

	Methodology	Analytics	Source
PSA	Data-driven	Descriptive	An EHR dataset
PRE	Data-driven	Predictive	An EHR dataset*
CPG	Knowldege-driven	Prescriptive	The NICE clinical guideline for Type 2 diabetes

Whereas, PSA [6] is a data-driven decision engine based on the patient similarity analytics that uses an EHR dataset from one of the largest hospitals in China and its affiliated community centers in order to manage type 2 diabetic patients. For example, given a patient encounter, the descriptive analytics of PSA reports that 75% similar patients take Metformin, while 15% similar patients take Insulin, and the rest take a combination of Insulin with oral anti-diabetic drugs. We employ various feature selection algorithms to identify the factors that affect physicians' prescription decisions. Given a patient encounter, his/her clinical conditions are represented using a vector of selected features, and we would find out the K most similar prescription instances, where the similarity is measured by the Euclidean distance between the representing feature vectors. Finally, PSA outputs a list of frequently presented medication options (among the K most similar prescription instances), and each option is attached with its occurrence percentage.

PRE is also a data-driven decision engine based on the outcome prediction analytics. The engine uses the same EHR dataset used by PSA but different features. For example, given a patient encounter, the predictive analytics of PRE reports that taking Insulin would get good outcome with support degree of 0.78, while taking Metformin would get good outcome with support degree of 0.54. Given a decision option, we collect the instances whose prescription is the same as the given decision option, and label its outcome as good or bad by comparing the next HbA1c test result after treatment with the current one. Thus, instances are grouped by different decision options, and we perform the feature selection (correlation based) and model training (logistic regression) per group separately. Then, given a patient encounter, for each decision option, his/her clinical conditions are represented using a vector of selected features, and we test it with the corresponding trained model. Finally, PRE outputs a list of decision options, and each option is attached with its support degree of good outcome.

Fusion engine

In the book [3], numerous methods for decision fusion have been presented. Based on the output of base decision engine, the fusion types are categorized into two: one is the fusion of label outputs, and the other is the fusion of value outputs. As described earlier, our base decision engines, CPG outputs labels, while PSA and PRE output values. It's easy to transform the CPG outputs as values, by assigning the label (which CPG recommended) with value of 1, and others (which CPG did not recommended) with value of 0, given a

patient at an encounter. Also, we can transform the PSA and PRE outputs as labels, but unavoidably, would have information loss. For instance, we could choose the most frequently presented medication option as the label for recommendation in PSA, and we could choose the decision option with the highest support degree as the label for recommendation in PRE. In this way we only keep the top one recommendation and ignore the lower ranked decision options – such a transformation from values to labels is not desirable. Consequently, our decision fusion framework is unified as a fusion of value outputs (and we will do a transformation from labels to values, if any base engine outputs labels).

Next, we will present a formal definition for our fusion engine. Suppose $E = \{e_1, ..., e_n\}$ be the set of base decision engines, $O = \{o_1, ..., o_m\}$ be the set of decision options, $v_{ij}(x)$ is the value that a decision engine e_j gives to the decision option o_i, for an instance x, $1<=i<=m$ and $1<=j<=n$. The fusion engine will take the following matrix $v(x)$ as an input, for each instance x.

$$v(x) = \begin{pmatrix} v_{1,1}(x) & ... & v_{1,n}(x) \\ ... & v_{i,j}(x) & ... \\ v_{m,1}(x) & ... & v_{m,n}(x) \end{pmatrix}$$

For fusion algorithm, we first apply the approach of decision templates [7]. At training phase, we calculate a decision template DT_i (as defined below) for each decision option o_i, where S_i is the set of training instances whose presecription is the same as the decision option o_i with good outcome, and N_i is the number of S_i. That is, DT_i is the mean of values of all training instances who take the decision option and get good outcome. Here, we highlight that S_i consists of training instances with good outcome, because our fusion is outcome-driven, instead of just learning the physicians' decisions (i.e. prescriptions, which do not always result in good outcomes).

$$DT_i = \frac{1}{N_i} \sum_{x \in S_i} v(x)$$

At testing phase, we calculate the squared Euclidean distance $d_i(x)$ between a decision template DT_i and $v(x)$ for an instance x, where $DT_i(k, j)$ is the $(k, j)^{th}$ entry in DT_i.

$$d_i(x) = \frac{1}{m \times n} \sum_{k=1}^{m} \sum_{j=1}^{n} (DT_i(k, j) - v_{k,j}(x))^2$$

The fusion engine outputs a list of decision options, and each option is attached with its distance value.

Besides decision templates, we also leverage various meta classifiers provided by Weka (a Java libarary for data mining [9]) to implement fusion algorithms. The main idea is to consider the values generated by base decision engines as new features, and feed them into a classifier for classification. Specifically, for each decision option o_i, we learn a model Mi, and the training data consists of instances x whose prescription is the same as the decision option o_i. The feature vector of x is $<v_{i1}(x), ..., v_{in}(x)>$, with label of 1 for good outcome and label of 0 for bad outcome, where $v_{ij}(x)$ is the value that a decision engine e_j gives to the decision option o_i, for 1<=j<=n. Thus, a variety of classifiers such as logistic and SVM (support vector machine) could be used for this meta-classification problem to decide whether the decision option o_i is a recommendation for an instance x.

Evaluation metrics

The motivation of decision fusion is "to do better" than base decision engines. This wish of "to do better" needs some evaluation metrics. As mentioned above, the often used metrics [4] such as precision, recall and RMSE are not applicable, when evaluating treatment recommendation systems, because we have only the partially observed ground truth – i.e., not all decision options are completely observed with outcomes. Formally, we denote $q(x) \in O$ as the prescription of an instance x, i.e., $q(x)$ is one of the decision option in O. Next, we denote $t(x, q(x))$ as the outcome of an instance x taking the prescription $q(x)$, where $t(x, q(x))=1$ means good outcome and $t(x, q(x))=0$ means bad outcome.

Back to the output of engines. All engines ouput values, but the values have different meanings. For example, PSA outputs the percentage of similar patients who takes the same decision option, and PRE outputs the support degree of good outcome which takes the given decision option, while the fusion engine outputs the distance between the decision template and the given decision option. Therefore, a value itself contributes little for evaluation, but a ranked list ordered by values does mean a lot. In particular, a rank score is calculated according to the position of prescription in a ranked list, which avoids to be overlooked when it's lowerly ranked.

For a formal representation, we denote $r_j(x) = <p_1, ..., p_m>$ as the ranked list of an instance x recommended by an engine $e_j \in E \cup \{e_0\}$ where each $p_k \in O$ is a decision option in O, and the subscript k means its position at the ranked list. Here, the base decision engine set E is union of the fusion engine e_0. We note that both PSA and PRE approach gets the recommended ranked list in descending order (because the more similar patients or the more supports, the better for recommendation), while our fusion algorithm of decision templates would get the recommended ranked list in ascending order (because the shorter distance, the better for recommendation).

Next, we denote $g_j(x)$ as the rank score of an instance x, given an engine $e_j \in E \cup \{e_0\}$. The following calculation means that, given a ranked list $r_j(x) = <p_1, ..., p_m>$ as recommended by an engine e_j, for any $0<=j<=n$, the prescription $q(x) = p_k$ is located at the position k, if an instance x taking the prescription $p(x)$ has good outcome, then $g_j(x)=(m-k)/(m-1)$, else $g_j(x)=(k-1)/(m-1)$. In particular, if the ranked list for recommendation does not contain the prescription, then its position is set as $k=m$. We note that, in spite of the partially observed ground truth (i.e. prescription with its outcome in real data), this calculation takes all recommended decision options into account, with information loss as little as possible.

$$g_j(x) = \begin{cases} \frac{m-k}{m-1} & if\ t(x, q(x)) = 1 \\ \frac{k-1}{m-1} & if\ t(x, q(x)) = 0 \end{cases}$$

$$where\ r_j(x) =< p_1, \ldots, p_m > and\ p_k = q(x)$$

The score of an engine e_j is calculated as $g_j = \frac{\sum_x g_j(x)}{\sum_x 1}$ for any $0<=j<=n$. Using this outcome-driven evaluation metric, we can directly compare the score g_0 of our fusion engine e_0 with other scores g_j of base decision engines $e_j \in E$.

Results

Our experimental dataset is an EHR dataset from one of the largest hospitals in China and its affiliated community centers that manage type 2 diabetic patients. After anonymity, it consists of 3150 encounter instances of diabetic patients. Their prescriptions are categoried as 7 types of decision options: METFORMIN (metformin alone), ARFA (either insulin secretagogues or α-glucosidase inhibitors), TZD (either thiazolidinediones or DPP-IV inhibitors), BI (two oral anti-diabetic drugs), TRI (three oral anti-diabetic drugs), INSULIN (insulin alone), and COMBINED (insulin and oral anti-diabetic drugs). For the outcome of glucose control, we use the widely adopted clinical ranges (also cited in [8]): HbA1c <= 6.4: normal; 6.5 <= HbA1c < 7: well controlled; 7 <= HbA1c<9: moderately controlled and 9 <= HbA1c: poorly controlled. Each patient's prescription outcome is labeled by comparing the next HbA1c test result after prescription with the current one. The outcome is labeled 1 (good) if the HbA1c level moves into a lower range, or remains in the well-controlled range, otherwise it is labeled as 0 (bad). In our dataset, there are 2574 instances of 3150 as labeled 1.

Given an instance, different base decision engines use different feature vectors as the input. Specially, CPG has 19 features, including overweight (BMI>24), old (age>70), etc. By feature selection, PSA identifies 39 features, such as average of glucose, value of HbA1c, and maximum value of HbA1c, etc. It's interesting that CPG considers the age status (e.g. old if age>70), while PSA considers the age (e.g. age of 72) and the age at the earliest diabetes diagnosis (e.g. age of 45). A more complicated case is PRE, which has 7 learning models for the 7 types of decision options. For example, the METFORMIN model in PRE has 13 features, while the ARFA model in PRE has 8 features.

In spite of different input features, the base decision engines have a uniform output format, i.e. v_{ij}, the value that a decision engine e_j gives to the decision option o_i, for an instance x, where e_1=CPG, e_2=PSA, e_3=PRE, and o_1=METFORMIN, o_2=ARFA, o_3=TZD, o_4=BI, o_5=TRI, o_6=INSULIN, o_7=COMBINED.

Regarding the fusion engine as e_0, it also follows the unified output format, and below is the sample output for an instance x. According to the real data, this instance x is prescribed o_2=ARFA, as having good outcome.

	CPG	*PSA*	*PRE*		*FUSION*
*o*1	1	0.4	0.6068		0.4286
*o*2	0	0.31	0.6943		0.2772
*o*3	0	0.01	0.7489		0.3196
*o*4	0	0.18	0.6580	⇒	0.5817
*o*5	0	0	0.5761		0.6454
*o*6	0	0.09	0.6674		0.4961
*o*7	0	0.01	0.5939		0.6164

The ranked list of x recommended by CPG is $r_1(x)=<o_1>$, PSA is $r_2(x)=<o_1,o_2,o_4,o_6,o_7,o_3,o_5>$ in descending order, PRE is $r_3(x)=<o_3,o_2,o_6,o_4,o_1,o_7,o_5>$ in descending order, and after fusion using algorithm of decision templates, it is $r_0(x)=<o_2,o_3,o_1,o_6,o_4,o_7,o_5>$ in ascending order.

The rank score of x given by CPG is $g_1(x)=(7-7)/(7-1)=0$, because $r_1(x)=<o_1>$ does not contain o_2, and its position is set as $k=m$, where $m = 7$ is the number of decision options. It makes sense, since CPG does not recommend o_2 which however results in good outcome in real data, and such a CPG recommendation gets the score of 0. For PSA, $g_2(x)=(7-2)/(7-1)=5/6$, because the position of o_2 in $r_2(x)$ is $k=2$ and $m=7$. It also makes sense, since o_1 is preferred by PSA than o_2, and such a PSA recommendation cannot get the full score of 1, but only 5/6. Similarly for PRE, $g_3(x)=(7-2)/(7-1)=5/6$. Here, we point out that, although PSA prefers o_1 while PRE prefers o_3 for the top one recommendation, the prescription o_2 is located at the same position $k=2$ in both PSA and PRE, so the rank scores $g_2(x)$ and $g_3(x)$ are the same. Finally, for fusion, $g_0(x)=(7-1)/(7-1)=1$, because the position of o_2 in $r_0(x)$ is $k=1$ and $m=7$. It matches the practice that such a fusion recommendation is appreciated.

Fig. 2 illustrates these treatment recommendation results, which could be integrated to the clinician workstation or mobile apps, and end users could have an overview about the engines' opinions in a comprehensive way.

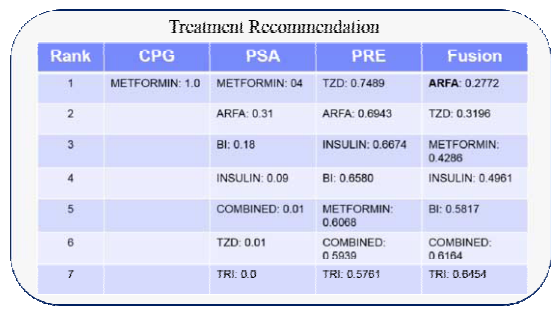


Figure 2 – Treatment recommendation results

To compare the evaluation results, we do 10-fold cross validation. That is, the total of 3150 instances is randomly partitioned into 10 equal size subsamples. Of the 10 subsamples, a single subsample (of 315 instances) is retained as the validation data for testing, and the remaining 9 subsamples (of 2835 instances) are used as training data.

Table 2 – Evaluation results

Treatment recommendation systems		Rank score
Decision engine	CPG	0.6771
	PSA	0.6815
	PRE	0.6062
Fusion engine	Decision templates	**0.6917**
	Logisic classifier	0.6826
	SVM classifier	0.6773
	SVM classifier –s 3	0.6823
	Naïve Bayes	0.6776
	Naïve Bayes -K	0.6824

As shown in Table 2, among base decision engines, PSA has a higher rank score (0.6815) than CPG (0.6771) and PRE (0.6062). After fusion, the engine of decision templates outforms all of the base decision engines, getting the highest rank score (0.6917). Also, the logistic classifier is promising and gets the rank score of 0.6826. Although we observe that

fusion engines using the default SVM classifier and Naïve Bayes have the rank scores lower than the base decision engine PSA, it's not a big drop since using machine learning algorithms always require careful tuning. Actually, we set an option "–K" for Naïve Bayes, which indicates to use kernel density estimator rather than normal distribution for numeric attributes, the rank score improves from 0.6776 to 0.6824. Similarly, we set an option "–s 3" for the SVM classifier, which indicates to use the SVM type of Epsilon-SVR rather than C-SVC, the rank score improves from 0.6773 to 0.6823.

Discussion

Fusion is not a new topic in machine learning and data mining, and its great success has been shown in a series of KDD Cup competitions from 2007 to 2014, as well as other open competitions such as the Neflix Grand Prize in 2009 [1] and the Heritage Health Prize in 2012 [2]. However, related work mainly focuses on the fusion of different learning algorithms with different parameter settings, and pays little attention to the knowledge sources. Furthermore, the evaluation metrics used in prior arts are based on the ground truth, but we have only the partially observed ground truth in treatment recommendation. To address these problems, we have two contributions as presented in this paper. First is a decision fusion framework for treatment recommendation systems, which combines both knowledge-driven and data-driven decision engines. Second is an outcome-driven evaluation metric, which has no information loss while facing the partially observed ground truth. As for experimental results, the fusion engine gets better performance than base decision engines.

Also, we realize our limitations. First, we assume that base decision engines output a uniform format. However, this assumption is challenged, if the labels (from different decision engines) are heterogenous. For example, suppose CPG just recommends using one of oral anti-diabetic drugs, but does not specify which one. Meanwhile, the decision options of PSA are still the 7 types: METFORMIN, ARFA, TZD, BI, TRI, INSULIN, and COMBINED. Now, one of oral anti-diabetic drugs could be METFORMIN, ARFA or TZD, how to assign values for such CPG recommendation? In our current work, CPG outputs label all the 7 types, and we do a simple transformation from these labels to the value of 1 or 0. Actually, we observe that normalization of heterogenous labels with values has not been mentioned or well addressed in previous work, because previous work is about multiple classifier systems, whose base engines are all classifiers to be trained for the homegenous labels. However, in our framework, we take both knowledge-driven and data-driven modules into account, which are heterogenous in nature. This situation would become more common, when facing the cloud-based decision services. We cannot assume each cloud-based decision service outputs a uniform format, but we still want to leverage the fusion of those services towards a better final decision. Therefore, the normalization for fusion would be our ongoing work.

Besides, in this paper, we only use three base decision engines of CPG, PSA and PRE for fusion. This is not enough, and we are planning to inclue more base decision engines involved towards an open (e.g. cloud-based) fusion platform. Moreover, we observe there is some contraindication information in the domain knowledge for treatment recommendation, e.g. you should not use statins for a woman becoming pregnant. Such contraindication information is hard to discover for learning algorithms, but it can be easily represented by knowledge modeling. Our future work will develop some novel fusion algorithm to get more benefits from both knowledge and data.

Last but not least, the so-called "partialness" of the ground truth deserves more investigation. For intervention in diabetes, we could regard the HbA1c value as partially observed, however, for a more nebulous topic, like antibiotics for fever, the partially observed ground truth may actually never have a known answer.

References

[1] The Netflix Prize. http://www.netflixprize.com/

[2] The Heritage Health Prize. https://www.heritagehealthprize.com/c/hhp

[3] Ludmila I. Kuncheva. Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience, 2004.

[4] Asela Gunawardana, Guy Shani. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. Journal of Machine Learning Research, 2009, pp. 2935-2962.

[5] Haifeng Liu, Jing Mei, Guotong Xie. Towards Collaborative Chronic Care Using a Clinical Guideline-Based Decision Support System, Proceedings of the 24th European Federation for Medical Informatics, MIE 2012, pp. 492-496.

[6] Haifeng Liu, Guo Tong Xie, Jing Mei, Weijia Shen, Wen Sun, Xiang Li. An Efficacy Driven Approach for Medication Recommendation in Type 2 Diabetes Treatment Using Data Mining Techniques, Proceedings of the 14th World Congress on Medical and Health Informatics, MedInfo 2013, pp. 1071.

[7] Ludmila I. Kuncheva, James C. Bezdek, Robert P. W. Duin. Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition, 2001, pp. 299-314.

[8] Hani Neuvirth, Michal Ozery-Flato, Jianying Hu, Jonathan Laserson, Martin S. Kohn, Shahram Ebadollahi, Michal Rosen-Zvi. Toward Personalized Care Management of Patients at Risk: The Diabetes Case Study. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, pp. 395-403.

[9] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 2009, Volume 11, Issue 1.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-305

Health Care Decision Support System for the Pediatric Emeregency Department Management

Sarah Ben Othman[a], Slim Hammadi[a], Alain Quilliot[b], Alain Martinot[c], Jean-Marie Renard[d]

[a]*Ecole Centrale de Lille, LAGIS UMR 8219, Lille, France*
[b]*University of Blaise Pascal, LIMOS UMR CNRS 6158, Clermont-Ferrand, France*
[c]*Pediatric Emergency Department "Jeanne de Flandre", Regional University Hospital of Lille, France*
[d]*University of Lille2, Lille, France*

Abstract

Health organization management is facing a high amount of complexity due to the inherent dynamics of the processes and the distributed organization of hospitals. It is therefore necessary for health care institutions to focus on this issue in order to deal with patients' requirements and satisfy their needs. The main objective of this study is to develop and implement a Decision Support System which can help physicians to better manage their organization, to anticipate the overcrowding feature, and to establish avoidance proposals for it. This work is a part of HOST project (Hospital: Optimization, Simulation, and Crowding Avoidance) of the French National Research Agency (ANR). It aims to optimize the functioning of the Pediatric Emergency Department characterized by stochastic arrivals of patients which leads to its overcrowding and services overload. Our study is a set of tools to smooth out patient flows, enhance care quality and minimize long waiting times and costs due to resources allocation. So we defined a decision aided tool based on Multi-agent Systems where actors negotiate and cooperate under some constraints in a dynamic environment. These entities which can be either physical agents representing real actors in the health care institution or software agents allowing the implementation of optimizing tools, cooperate to satisfy the demands of patients while respecting emergency degrees. This paper is concerned with agents' negotiation. It proposes a new approach for multi-skill tasks scheduling based on interactions between agents.

Keywords:

Decision Support System; Overcrowding; Pediatric Emergency Department; Multi-agent Systems; Physical Agents; Software Agents; Negotiation; Multi-skill tasks scheduling.

Introduction

The rapid technological advances in computer science, telecommunication, and networking fields have led to the birth of a new paradigm for health care field referred to as eHealth; and also the integration of modern networking technologies in the medical process. In fact, in the last decades the link between engineering and medicine has become closer. Intelligent systems are created in order to automatically process medical data and provide support for medical decisions [1]. Information Technologies (IT) have provided major breakthroughs over the last several decades. Breakthroughs in health care emergency structures which have decided to work out an information system of health control supplied by the emergency services; and the establishment of internal information systems supporting the pre-hospital activities and downstream emergency services. An effective information system is considered by emergency professionals as a major element in the efficient functioning of the emergency structure. It should allow patient flow and medical records management, knowledge of availability of medical staff members, and a qualitative and quantitative knowledge of the service activity [2]. This information system can also promote more effective management of health care systems (dashboards, personalized program information system, summary measure of activity, computerization of medical practices, coding acts, etc.). Health care systems are being developed to provide support for medical decisions using health information networks; and portable communicable systems combined with many other tools and applications based on Information and Communications Technology (ICT) in order to assist physicians/medical experts in patients' treatment and health monitoring and management [3, 4].

Our collaboration with the Regional University Hospital (RUH) of Lille has led to the HOST ANR national project. This project is a scientific research work aimed at health care field with fallouts for the Pediatric Emergency Department (PED) of Lille RUH. The ubiquity of the emergency problem in France and in Europe, the strategic place of the PED in health establishments, patients' dissatisfaction, and rising costs of health care services are primarily responsible for the high level of our interest in PED. We aim through this project to study PED modeling, optimization of its functioning, and development and implementation of a Decision Support System (DSS). This is to anticipate overcrowding feature in the hospital and to establish avoidance proposals for it.

In this paper, we propose a distributed architecture designed for health care organization management, where actors provide smart negotiation in order to execute and control health care treatment tasks. The presented solution comprises use of an agent-based framework including an optimization and negotiation scheme to solve multi-skill health care tasks scheduling.

Methods

Multi-agent system architecture

Several models exist to design a dynamic system with a collective behavior. Since patients' health care in emergency department is a complex problem that requires a good division of tasks and data, a Multi Agent System (MAS) is used to model the PED organization. It is capable of identifying the

main functions of the system, as well as tasks, goals, and missions. Intelligent agents are therefore a good option for designing a decision support system for emergencies services [5, 6].

The literature on health care logistics shows that distributed decision-making agent-based paradigm provides an interesting approach to improve quality of care services provided by health care facilities. The Telemedicine-Oriented Medical Assistant (TOMAS) is a health research project developed using agent-based architecture. It is used by each specialist to transfer microscopic images and patient data in order to develop a diagnostic in collaboration with other colleagues in pathology department [7]. The Intelligent Healthcare Knowledge Assistant (IHKA) is an intelligent assistant in health care field using MAS for dynamic collection of knowledge, filtering, adapting, and care units acquiring [2]. Zachewitz developed an agent-based system to improve immunization rates in Germany [8]. Agents have been developed to maintain the consistency of patient's medical file at the pharmacy, at the level of specialists, and assigned doctors.

To improve logistics planning, we propose a dynamic multi-agent architecture (Figure 1).

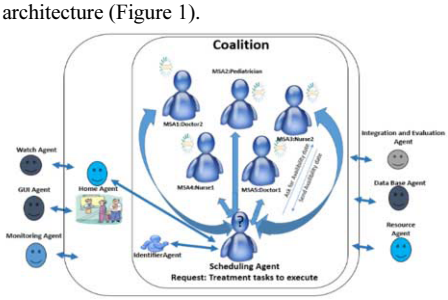


Figure 1- The proposed agent based architecture

It considers each actor in the PED as an autonomous agent capable of exchanging information with other actors in a self-sufficient way [9]. An agent can represent a real physical entity involved in the system as patient, medical staff, etc. or functional entity to resolve to solve complex problems such as scheduling, identification, etc. Autonomous agents and MAS have led to a new way to view, analyze, model, and design complex software systems.

Physical agents

Physical agents represent the parts which are really acting in the PED. These include different members of the medical staff charged to treat patients, reception nurses, orientation nurses, etc. In the proposed architecture we have:

Home Agent (HA): It represents activities of the reception nurse. When a patient arrives to the PED, it is responsible for receiving patients, their orientation, and for process pathology identification. HA has all skills required by the rules of registration plan, medical diagnosis plan, and patient orientation plan. HA deals with the problem formulation and then sends it to the Identifier Agent (IA) corresponding to the creation of a medical record through the IA triggered by an administrative nurse.

Identifier Agent (IA): It models the orientation nurse activity. It receives different information from HA about the medical problem and identifies skills needed for treatment referring to medical protocols. It consults the database of different pathologies and needed resources for patients' treatment.

Medical Staff Agent (MSA): MSA models medical staff activities. It is a mobile software agent which can move intelligently from one medical team to another in the PED in order to treat patients. It is characterized by two variables (skills and availability). In our proposed system, we use mobile medical staff agents to travel through the PED architecture to treat patients according to their availabilities and to the emergency degree of patients; and to collect intelligently, needed information related to patient health state in order to update the system data. This special type of agent has a smart behavior and is composed of data, states, and a code. Once MSA achieves a treatment task, it can shift to another medical team for new task execution. Therefore, while scheduling we must take into account this aspect when assigning human resources to tasks. Each task represents a service which can be performed by different possible MSAs, with different cost.

Software agents

These agents have the role to insure functioning of proposed systems. They allow development and implementation of various optimizing tools, different interactions, and communication protocols. The software agents in our system include:

Scheduling Agent (SA): This agent is responsible for optimizing choice of resources for patients' treatment taking into account our system constraints. It has to assign resources to patients' treatment tasks minimizing total cost and patients waiting time in order to respect emergency degrees. It organizes queue of patients who need treatment taking into account their emergency degree. It then assigns resources to different tasks.

Integration and Evaluation Agent (IEA): This agent is responsible for the whole system performance control. It calculates performance indicators of the system such as waiting time of patients and treatment costs in order to evaluate overall schedule of patients in PED.

GUI Agent (GA): This agent interacts with system users particularly medical staff in PED.

Resource Agent (RA): This agent is responsible for resources monitoring.

Monitoring Agent (MoA): This agent is notified of every decision taken and every task completed. It represents the coordinator between all software agents; and an informer for the physical agents about actions and patients status.

Scheduling strategy: Negotiation protocols

The coordination of distributed planning approach focuses on interactions between agents' plans. An interaction protocol allows agents to have conversations in the form of structured exchanges of messages [10]. In an emergency service, medical staff members are often faced with treating multiple patients simultaneously. Thus, a patient can be treated by various doctors and nurses. Consequently, medical staff in the PED is organized in a cooperative group that shares same resources, that is to say, patients. The medical staff of the PED is working on basis of dynamically generated plans by SA which contains details of treatment tasks operations to execute. This coordination is achieved by appropriate protocols of interaction between agents.

To carry out scheduling, SA needs information about medical staff availabilities to execute treatment tasks. SA sends a request to MSA for execution of processing task and negotiations between SA and various mobile agents modelling medical staff members occur.

The SA is the initiator of the negotiation. It ignores relevant information to various resources to be allocated to patients'

treatment tasks, but it knows all information about patients, their pathologies, and their requests.

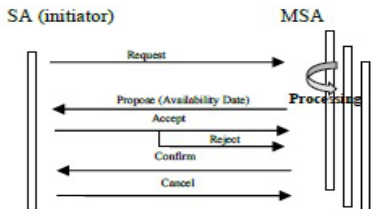


Figure 2 - Negotiation Protocol

The initiator (SA) sends a request to MSA deemed qualified to execute health care operations. The message form is as follows:

<SA, MSA_k, Request, AvailabilityDate (k), $V_m(t)$>, where ;

AvailabilityDate (k) is earliest time at which medical staff member MS_k is able to execute the set of treatment tasks $V_m(t)$.

Vm (t) is set of operations to be treated by MS_k on different patients at time t.

When a MSA receives a query, it analyses, calculates in real time its availability date, and sends response to SA which makes a comparison between availability dates provided by all MSA then decides (accept or reject proposal). The MSA with the earliest availability date is chosen. The availability date of each medical staff member depends on his skills as well as current medical treatment task evolution.

Simulation and Results

Our partner is the PED "Jeanne de Flandre" of RUH of Lille (North France). We noted 47,188 visits during 2011-2012 ranging from neonates to 18 years. 19% of them were hospitalized in Short-stay Observation Unit while 81% of the patients were of Outpatient Care Unit. We focused our study on the second population, the source of crowding. In fact, length of stay usually does not exceed 8 hours otherwise patient is transferred to Short-stay Observation Unit.

System implementation

Our system is developed with JADE (Java Agent DEvelopment framework) platform [11]. JADE simplifies implementation of MAS through a middleware that complies with FIPA (Foundation for Intelligent Physical Agents) specifications and provides a set of graphical tools supporting debugging and deployment phases. Figure 3 shows evolution of message exchange between the different agents through "sniffer" tool useful for debugging.

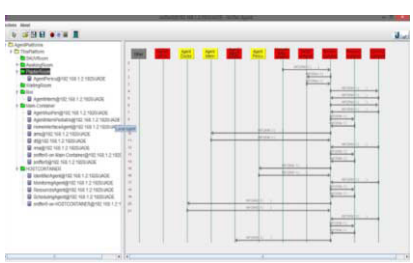


Figure 3 - Messages exchange

JADE system supports coordination between several agents and provides a standard implementation of communication language FIPA-ACL, which facilitates communication between agents which complies with FIPA specifications [12]. JADE is written in java language. It supports mobility. In this paper, we used a JADE graphical tool which sniffs message exchange between agents. This tool is useful to debug a conversation between agents.

All the system is implemented as a decision support tool; software based on a MAS. So, we programmed scheduling approach and we have integrated different calculation formulas. It is a flexible tool where we can adjust and act directly through ergonomic interfaces.

To better explain our approach solving the discussed scheduling and optimization problem; and functioning of our system, we propose the following illustrative example showing different aspects and logic followed by our proposed scheduling system. We start with the system up and running in PED department with patients already signed up with registration nurse. The initial situation is thus with medical staff already treating patients. We start from $t=t_0$, where a new patient arrives at the PED reception. Patients' information is requested by nurse and processed into our system in order to make new patient's data available to HA. Data received by HA will then pass through a pathology test in order to collect various data including emergency degree, pathologies, and needed resources.

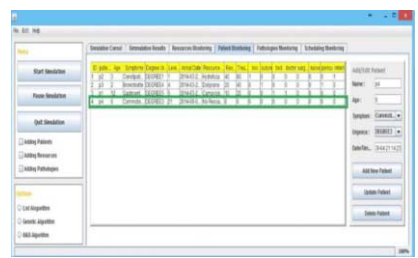


Figure 4 - Adding a new patient

These data will be used to generate a priority degree affected to the newly arrived patient giving him a position in the waiting queue.

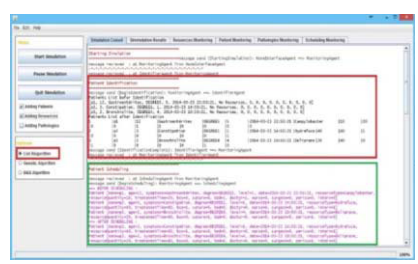


Figure 5 - Patients' identification and scheduling

The position is generated using List Algorithm with dynamic and flexible priority rules. Therefore, all patients' position will change according to the new scheduling in order to satisfy all patients in a matter of emergencies and waiting time.

As a result we will have an initial scheduling (Figure 6) consists of 9 care operations, for example, assigned to two medical staff "nurse" and "doctor" able to execute them. Medical care procedures can be done at the same time (for example, in case of concussion, a doctor makes a diagnosis

while a nurse is doing a careful neurological exam for the same patient). Therefore, operation can be performed carefully by mobilizing members of Medical Staff at same time and with same duration; or with varying execution times according to the skills required for the treatment realization.

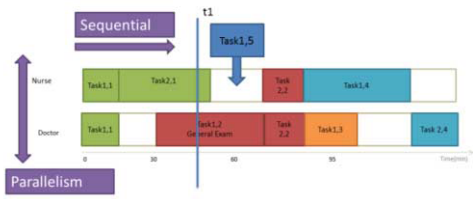


Figure 6 - Tasks scheduling

The affectation of medical staff is based on needed skills and on availabilities. Therefore, we use a real time and dynamic Time Table algorithm for human resources scheduling aiming to reduce idle time which leads to costs minimization. This scheduling takes advantage of agents' mobility to optimize patients' treating and waiting time; and also to optimize medical staff operating time which leads to cost minimization.

We always start by treating the most urgent request using available resources. For human resources, if the member of medical staff has needed skills to treat a treatment request, the most urgent treatment task is allocated to him. The decisions are the result of coordination and negotiation between our system's agents and are taken in a dynamic manner.

At $t=t_1$, a new patient arrives. A doctor is needed for his treatment. Thus, the doctor (MSA) leaves the box where patient 1 is being treated to go to another box to treat patient 5. An operation may also be accomplished or interrupted by an emergency.

The graph in Figure 7 represents evolution of patients' pathways and health care activities being completed for each patient. The parallel line to x-axis represents time spent in PED.

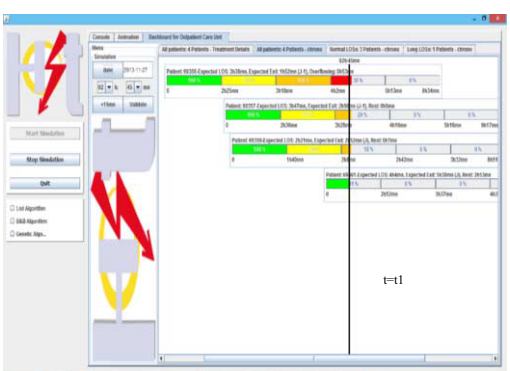


Figure 7 - Patients tracking

For example we have in figure 7 above 4 patients; the first one has arrived first, then patient 2, patient 3, and 4. At $t=t_1$, length of stay of first patient is the longest one, so he needs attention. However, the last patient has already arrived, so if his emergency degree is not high, we consider that he undergoes normal treatment process and there is no need to worry about him at the moment. This tracking helps to know at each moment how many patients we have as well as evolution of their treatment process. It also improves service quality and information delivery through indicating length of stay of each patient in the PED.

Simulation scenario

To characterize our model, we used the 3-year data collected in the PED of CHRU of Lille. We have information about 23,150 patients' admissions in 2011 and 24,039 in 2012 (Figure 8). In order to protect individual privacy, patients' data were previously strictly anonymized.

Figure 8 - 3-year data of CHRU of Lille PED

We considered the scenario of November 28th, 2013 because it presents the highest level of crowding throughout the first 11 months of 2013 according to physicians. So, we will use it as simulation scenario for our results validation. At 08:00, we had two patients, then arrivals gradually increased until reaching a first peak of 17 patients at 11:15, then three hours later, a second peak of 27 patients at 14:15.

Figure 9 shows that the proposed decision support system allows smoothening of PED activities and absorption of large demand load. Indeed, use of the discussed methods reduces peaks of activity and minimizes patients waiting time. The green curve represents PED activities without using the decision support system and the black curve shows optimization effect.

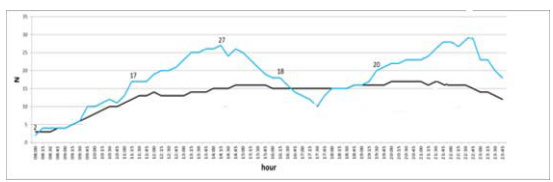


Figure 9 - PED activities smoothing

Medical Staff Members Activities

MSA in the PED, because of their different properties (autonomy, communication and interactions, cooperation, etc.) may have different features of activities. Indeed, at a given time t, some agents negotiate between them, others transmit commands while others are executing scheduling tasks. During an observation in the PED, we identified four modes of activity, according to Figure 10;

- Parallel Activities: In this example, the four proposed agents are working simultaneously, non-sequentially. Agent waits at end of an activity to begin another.
- Staggered Activities: All of the agents work with a lag. An agent can start a new activity even if a previous activity is still being executed.

- Interruption: A MSA that can stop an activity already begun, if it receives an update or a more urgent task to execute.
- Different durations: The activities of agents are of different durations. These durations depend on type of activity being carried out.

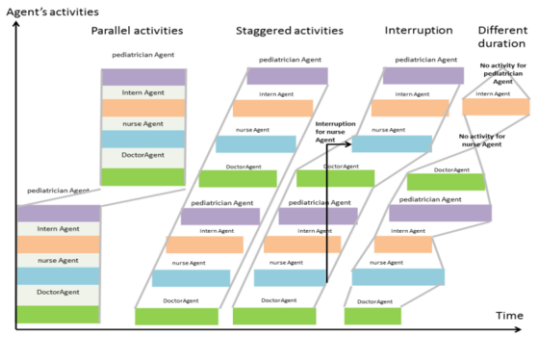


Figure 10 - Agents activities

Discussion

During our study, we noticed significant stress upon medical staff members; and dissatisfaction of some patients because of long waiting times. We have also noticed difficulties in management of patients especially in cases of an overcrowded PED. So, health care specialists should get involved in new actions aiming to improve care services. Requirements to satisfy patients' requests impose need to adopt new approaches to solve problems related to patients' treatment process management; and human and material resources allocation.

Our primary consideration in this work is development of decision support solutions to simplify the work of medical staff in the PED. From monitoring screens, medical staff can have complete and accurate feedback of the global state of the PED. The proposed software is capable of determining medical staff availabilities which are uncertain and depend on the experience level of the health care provider; and evolution of the medical act being executed. Because time is important in emergency services, poor control of availability of professionals can generate a worsening of the condition and cause of preventable care. Our solution for rapid and effective treatment of patients is based on alliance between MAS paradigm and optimization tools; and improving use of human resources through cooperation of different members of the medical staff; and dynamic and reactive generation of treatment plans. The use of intelligent agents in medical field has been demonstrated as a complementary technique to improve performance of computer. Traditional centralized control mechanisms show insufficient flexibility to deal with sudden changes of needs. In fact, in the PED, information circulates poorly between medical staff members and is sometimes missing. So, to ensure real time negotiation between the different actors of PED, MAS is the most suited tool for real time interaction and negotiation. MAS modeling consists of modeling agents with their activities and interactions.

Conclusion

In this paper, we have proposed a distributed scheduling system for care of patients in the PED. This system can contribute to health care quality improvement, costs minimizing and waiting time reduction. Our research has therefore focused on definition of an agent-based approach to execute patient treatment tasks. This application displays utility of a decision support tool which can help to manage patients' treatment in the PED. For future work, we believe that there is need for further improvement of the scheduling algorithm to be integrated in the system.

References

[1] Chiuchisan I, Geman O. An approach of a decision support and home monitoring system for patients with neurological decision makers using internet of things concepts. WSEAS Transaction on Systems 2014; 13: 460-469.

[2] Tan J. E-health care information systems: an introduction for students and professionals, Jossey-Bass Press, 2005.

[3] Pepe G, Frumento E, Garlaschelli A. A service for remote screening of physiological data and follow up of patients using a home gateway and domestic wireless technologies. In: Bos L, Laxminarayan S, and Marsh A, eds. Medical and Care Compunetics 1. Amsterdam: IOSPress, 2004; pp. 299-306.

[4] Tchernev N. The modeling of logistics process in flexible production systems. 1997. Ph.D in Computer Science dissertation, Blaise Pascal University, Clermont-Ferrand.

[5] Bartelt A, Lamersdorf W, Paulussen TO, Heinzl A. Agent oriented specification for patient-scheduling systems in hospitals. Dagstuhl Article, 2002; 5: 1-8.

[6] Paulussen TO, Jennings NR, Decker KS, Heinzl A. Distributed patient scheduling in hospitals. In: Proceeding of the 18th International Joint Conference on Artificial Intelligence; Acapulco, Mexico. Morgan Kaufman publishers Inc.; 2003: pp.1224-1229.

[7] Della Mea V. Agents acting and moving in healthcare scenario: a paradigm for telemedical collaboration. IEEE Trans Inf Technol Biomed 2001; 5(1): 10-13.

[8] Zachewitz L. A multi-agent system for establishing electronic healthcare records. In: 16th European Conference on Artificial Intelligence (ECAI 2004) - Workshop 7: Second workshop on Agents Applied in Health Care; August 2004; Valencia, Spain. 2004. pp. 31-37.

[9] van der Linden H, Kalra D, Hasman A, Talmon J. Inter-organizational future proof EHR systems: a review of the security and privacy related issues. Int J Med Inform 2009; 78 (3): 141-160.

[10] Huhns MN, Stephens LM. Multi-agent systems and societies of agents. In: Multiagent systems. Cambridge, Massachusetts. MIT Press; 1999. pp. 79-120.

[11] http://jade.titlab.com/doc

[12] www.fipa.org

Address for correspondence

Sarah Ben Othman, LAGIS UMR CNRS 8219, Ecole Centrale of Lille, Avenue Paul Langevin, Cité scientifique, 59650 Villeneuve d'Ascq, France, e-mail: ben.othman.sara@hotmail.fr.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-310

Analysis of Hospital Processes with Process Mining Techniques

Arturo Orellana García[a], Damián Pérez Alfonso[b], Osvaldo Ulises Larrea Armenteros[c]

[a] *Department of Applications, Center of Medical Informatics, University of Informatics Sciences, Havana, Cuba*
[b] *Professor at the faculty 3, University of Informatics Sciences, Havana, Cuba*
[c] *Engineering student, University of Informatics Sciences, Havana, Cuba*

Abstract

Process mining allows for discovery, monitoring, and improving processes identified in information systems from their event logs. In hospital environments, process analysis has been a crucial factor for cost reduction, control and proper use of resources, better patient care, and achieving service excellence. This paper presents a new component for event logs generation in the Hospital Information System or HIS, developed at University of Informatics Sciences. The event logs obtained are used for analysis of hospital processes with process mining techniques. The proposed solution intends to achieve the generation of event logs in the system with high quality. The performed analyses allowed for redefining functions in the system and proposed proper flow of information. The study exposed the need to incorporate process mining techniques in hospital systems to analyze the processes execution. Moreover, we illustrate its application for making clinical and administrative decisions for the management of hospital activities.

Keywords:

Hospital environment; Process mining; Event logs; Hospital Information System; Process model.

Introduction

Demand for medical services at an international level is increasing. The costs related to search and implementation of strategies and tools to meet the medical needs of populations push for an improvement in the execution of clinical and administrative tasks in health institutions. In order to address this situation, hospitals need to find alternatives that make their processes more efficient. The improvement in quality of processes implies higher productivity [1], so the design of new methods to optimize performance of processes is necessary.

The business process management approach allows for identification and management of processes that are interrelated; in order to analyze organizational performance and make continuous improvements of the results, eliminating errors, and redundant processes in the institution. Process mining allows for understanding how processes are really executed in the system. Its application helps to identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures, and simplify processes [2] for improving business performance. This technology allows getting information about how the process has behaved, the times it takes in the executions, as well as variations between reality execution and the prescribed model. The complexity of this approach is proportional to the amount of data stored; making it invaluable information for decision-making. These data are usually stored in the databases of the systems that support organizational processes. From these databases, certain information could be extracted to create event logs.

An event log is the evidence of events that occur in the computer systems and the networks of an organization [3]. The records are composed of input events. Each entry contains related information to an action or specific operation that was executed on the information system and its corresponding timestamp. These event logs should be formatted as MXML [6] or XES [7].

Currently, to obtain event logs from information systems, it is necessary to use specialized tools like: XESame [4] and Nitro [5]. Even though these software are efficient, their high complexity related to technical configurations and database structures make them inconvenient for inexperienced or non-expert users.

With the aim to computerize, the processes of secondary healthcare level have been developed into a Hospital Information System (HIS) at the University of Informatics Sciences (UCI), in the Center for Medical Informatics (CESIM). This HIS supports the fundamental processes carried out at this level of healthcare and was designed to meet the needs of: storage, processing, gathering, and interpretation of the medical-administrative data generated.

The HIS from UCI does not have a component to generate event logs from the execution traces stored in the system; thus precluding application of process mining techniques for analysis of running processes. Medical and administrative staff have limited access to this source of knowledge for clinical and administrative decision-making.

Due to the above inconveniences, we propose to develop a software component that allows generation of event logs from the Hospital Information System without relying on other external tools. This system component generates event logs with the same characteristics, format, and structure as obtained with XESame tool; but can be used by non-expert users. This solution allows analysis of healthcare process using process mining techniques.

This article has two main sections. The first deals with implementation of a software component to extract traces from the Hospital Information System. The second section focuses on process analysis with process mining techniques, starting from an event log generated using the developed component.

Materials and Methods

Processes mining in a hospital environment

An intelligent approach to healthcare is use of information to create knowledge about patient care and improve performance of the hospital institution. Healthcare organizations increasingly have more data, which involves a large amount of information and few alternatives for analysis; a result of the lack of effective tools or methods that allow for obtaining the decision criteria about execution of the system processes.

The process mining techniques allow extraction of useful and nontrivial information from execution logs stored in information systems [8]. There are three types of process mining techniques: discovery, conformance checking, and enhancement or extension. All these techniques allow us to analyze actual execution of processes. The discovery technique is used to extract process models from an event log. Conformance checking techniques are used to monitor deviations by comparing the model and the event log and improving existing models from its extension [9].

Investigations carried out provided evidence that process mining is a novel and effective technology for analysis of hospital processes. In Gynecology and Oncology [10], process mining techniques helped to optimize, from dotted points, the trajectory of patients through the process of care. Also, discovering the busiest times in the area of Emergency [11] allowed for control and allocation of resources to this hospital sector. Other research in the area of Emergency [12] presented the detection of eventualities (incomplete assignments, missing information, little correspondence between the business process and system) in the process activities using process mining techniques. Its application in the Electronic History Records [13] allowed for enhancement of patient care processes.

Maruster [14] proposed the application of Petri Nets based on Process Mining for modeling hospital information. Moreover, the application of process mining on hospital systems allowed monitoring use of resources such as surgical implements, for nursing and external consultation. Process mining applications in other environments [15-18] demonstrate its effectiveness in detecting malfunctions, such as deviation of resources, extra stay time of a product in a warehouse stock or loss of any product, in such a way that allows for adjustment of product ordering frequency. From the engineering point of view, programming errors in the processes definition and activities that do not run properly or are simplified could be identified. Process mining techniques allows finding bottlenecks in information flow.

Event Log

To formalize the structure of event logs used in process mining, two standards were defined: Mining eXtensible Markup Language (MXML) and eXtensible Event Stream (XES). MXML is a format based on XML for the exchange of event logs. MXML was the first standard that emerged in 2003 and was adopted by the process mining tool ProM [19].

MXML establishes a standard notation to store dates, resources and types of transactions. In 2010, XES replaced MXML as the new process mining format independent of the tool [17]. XES is a standard based on practical experiences of MXML, and is less restrictive and truly extensible. Its main purpose is to provide an interchange format for event logs between tools and application domains [7].

To ensure a successful analysis of process mining, the quality of the storage format of the event log should be ensured. This is defined from three fundamental aspects: reliability, completeness, and security. Reliability stands for that ability to safely assume that the recorded events really occurred and that the attributes of the events are correct. Completeness is the characteristic of not missing any event in a certain context. In addition, any registered event must have well-defined semantics. The event data are considered safe when considerations of privacy and security are taken on, when events are registered [18].

Attributes of Event Log

An event log contains evidence of processes execution stored like traces. This record of elements can also contain attributes. As the registration of elements is created only once, the impact of including many attributes in the registry is minimal [6]. It is important to include relevant information describing the contents and origin of the event log. The following attributes should be taken into consideration for trace recording:

- **Process Name**: The name of the process under which the log will record its execution.
- **Data Source**: A description of the information system from which the event log is extracted from.
- **Source Organization**: The name of the organization providing the data.
- **Description**: A brief description of the contents of the event log.
- **Version**: An identifier to differentiate versions of event logs.
- **Author**: Name and contact details of who defined the conversion.
- **Process Mining Project**: A reference to the Process Mining Project or purpose of the event log.

Development environment

We proceeded with the implementation of the software component from the theoretical study of event logs. The following technologies and tools were used:

Eclipse Ganymede 3.4.2: Eclipse is an IDE (Integrated Development Environment) open and multiplatform code that has achieved a high degree of maturity in the development of what are known as "rich client applications". It has tools to develop console applications and web services with different application servers such as JBoss, Websphere, and Glassfish. It was originally developed by IBM (International Business Machines) and its future is now in the hands of Eclipse Foundation, an independent nonprofit organization that fosters an open source community and a set of complementary products, capabilities and services.

JBoss Application Server 4.2.2: JBoss Application Server is a Java EE (Java Enterprise Edition) free software implemented in pure Java. Being based in Java, it can be used on any operating system that supports it. It provides a full range of services for Java EE 5 and expansion of business services, including clustering, caching, and persistence. JBoss is ideal for Java applications and applications based on the web. It also supports EJB 3.0 (Enterprise Java), making applications development much simpler.

JBoss Seam V2.1.1: JBoss Seam is a powerful framework for developing Web 2.0 applications by unifying and integrating technologies such as AJAX, JSF, EJB, Java Portlets and BPM (Business Process Management). Another important feature is that POJOs (Plain Old Java Objects) can be validated in addition to directly handling the application logic and business sessions from beans.

Java V1.6: Java is an object oriented programming language developed by Sun Microsystems in the early 1990s. The language itself borrows much of its syntax from C and C ++ but has a simpler object model and eliminates tools of low level, which can cause many errors such as direct manipulation of pointers or memory.

Hibernate V3.3: Hibernate is an ORM (Object Relational Mapping) tool for Java platform. It facilitates the mapping of attributes between a traditional relational database and the object model of an application using declarative XML files or annotations, on the beans of entities that allow for establishing these relationships.

Facelets V1.1: Facelets is a simplified presentation framework where it is possible to freely design a web page and then associate specific JSF components. It provides more freedom to the designer and improves error reporting, having JSF. It allows for definition of disposition of pages based on a template, the composition of components, creating custom labels, and facilitates friendly development for the graphic designer and creation of component libraries.

Results

Figure 1 shows the user interface for the component developed to generate event logs in XES format. This component requires a process name as input, specified by the user to identify the process. It is needed to select the process to be analyzed and delimits start and end date for the events that will be inputted in the event log.

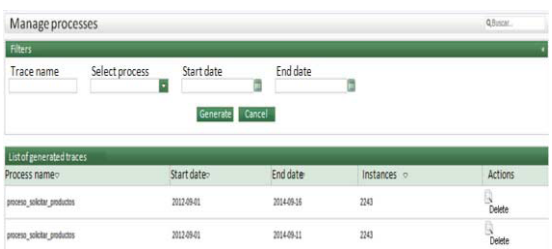


Figure 1 - Component to generate an event log in the HIS.

Below is an excerpt of the event log (see Figure. 2) obtained from the component developed. The extensions and attributes mentioned above are reflected in this image.

```
<trace>
        <string key="concept:name" value="4364"/>
        <string key="description" value="Simulated process instance"/>
        <event>
                <string key="org:resource" value="cirujano"/>
                <date key="time:timestamp" value="2012-01-15T16:53:24.491.000+01:00"/>
                <string key="concept:name" value="ver_detalles_sol_bq"/>
                <string key="lifecycle:transition" value="complete"/>
        </event>
        <event>
                <string key="org:resource" value="administrador"/>
                <date key="time:timestamp" value="2012-01-15T16:53:35.813.000+01:00"/>
                <string key="concept:name" value="despacho_sol_bq"/>
                <string key="lifecycle:transition" value="complete"/>
        </event>
</trace>
```

Figure 2 – Event log excerpt.

An event log related to the process: Request Product of HIS Warehouse module, was obtained from the developed component. The Warehouse module, among other activities, manages the flow of information on the various movements that a product can have in a warehouse. There are three types of product requests: tendering request, application warehouse, and request for surgical block (see Figure 3). To show the usefulness of the developed component, the Request Product process from Warehouse Module of the Hospital Information System created at UCI, was selected. An event log related to this process was extracted and analyzed with the ProM tool as shown below.

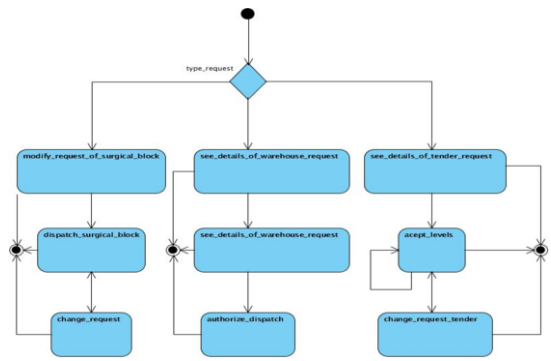


Figure 3 – Business Process Diagram of Request Product.

This event log was loaded on the ProM tool in order to conduct the process mining analysis. The process model showed on Figure 4 was obtained by applying the technique “Mine Net using Heuristic Miner”. This process model is a representation of the actual process behavior recorded in the event log during its executions on the HIS.

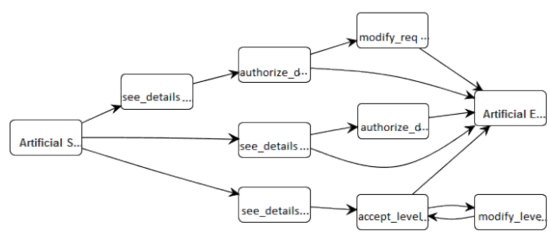


Figure 4 – Model obtained by applying the technique Mine for a Heuristic Net using Heuristic Miner.

This technique shows the similarities between the formal definition of the process and the event log obtained from the traces of their execution. In this model, there exists a difference with respect to the reference model; the process never ends directly after see_details_tender_request activity and whether it should end.

In comparing Figures 3 and 4, the applied technique shows that the process execution in the system is in concordance with the formal definition of the business process. The system has been implemented in the way it its business was conceived.

Figure 5 illustrates the observed result of the technique: "Analyze using Dotted Chart". The Dotted Chart shows an overview of the process where the X-axis shows the time, the Y-

axis shows the process execution instances and each point represents an event; the colors of the points refer to the different activities.

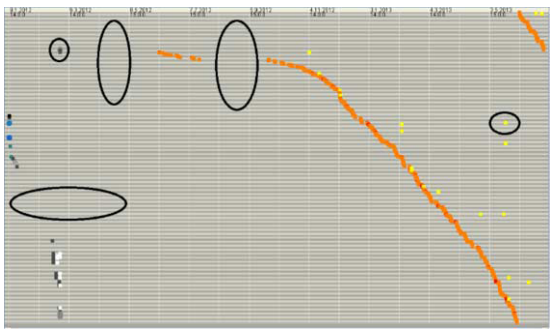


Figure 5 – Aplication of technique "Analize using Dotted Chart" whit black circles representing noise.

The runtime process was extended over seventeen months and three distinct moments with different behaviors could be distinguished. The first moment corresponds to system testing stage, where see_details_of_warehouse_request and authorize_dispatch activities were not executed. Afterwards, the system does not record any execution of the process for three months. The second and third phase were extended a month and a half to nine and a half months, respectively, with a downtime of two months between them; only see_details_of_warehouse_request and authorize_dispatch activities were running. Because of these characteristics, this process shows an anomalous behavior as 6 of the 8 activities making up the final version of the process are left to run.

The next model (Figure 6) represent the technique Replay a log on Petri net for Performance/Conformance, that allows some cases to define in a simple way, for example, the more common activities flow that runs in the system process under analysis.

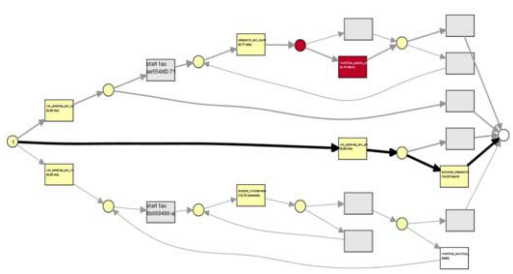


Figure 6 – Aplication of technique "Replay a log on Petri net for Performance/Conformance".

This technique establishes that while darker and thicker lines in the figure indicate a transition, a greater amount of times have been carried out for the activity represented in the flow.

This model allows identification of the flow of the more executed activities in the processes under analysis. Therefore, in the process Product Request, the flow with more executed activities can be defined as composed by the activities see_details_of_warehouse_request and authorize_dispatch.

In comparing the difference of maximum and minimum waiting times between modify_request_of_surgical_block and authorize_dispatch activities, we see that:

- The minimum waiting time of the first activity (3.29 seconds) is less than the second one (4.43 minutes).
- The timeout of the first activity (5 months) is greater than the second activity (7.45 days).
- The average time of the first activity (14.38 hours) is less than the second activity (3.31 days).

These analyses allowed receipt and dispatch of products from the warehouse, as well as detection of a bottleneck in the activity authorize_dispatch to be controlled. The understanding and correct interpretation of the process models obtained from process mining techniques have been discussed in the literature, as challenges [19] to analyze the behavior of computer systems by non-experts in the field of knowledge are the same.

Conclusions and future works

Process mining is presented as an effective alternative to show a radiograph of the actual process execution. However its models are far from favorable to non-experts.

The application of process mining techniques in the hospital environment is novel because there are only a few cases of its study around the world. However there is support for its use in this key social sector.

We developed a component for extracting traces of the HIS, which overcomes the problems raised above. From the analysis of the traces generated by the developed component using process mining techniques, we concluded that they contain valuable information useful for process analysis.

The solution developed guarantees that data persist permanently in the trace processes records. As a part of the developed component, it ensures transformation of stored records in the HIS to processes event logs. It also allows to export the to XES format.

After applying multiple modeling techniques to the event log obtained, was stated that the obtained models are not suitable for understanding by health personnel. The incorporation of process mining techniques to the HIS where models are generated, allowing for analysis by non-experts, is plotted as a goal of this research. Towards this goal, we conduct research on what elements are essential to the analysis and understanding of process execution.

The component developed is a starting point for incorporation of process mining techniques and algorithms of the HIS, in order to support clinical and administrative decision-making. This component becomes the first of its type, designed to analyze processes in the systems HIS.

Understanding the process model (in terms of accuracy and efficiency of understanding) is a function of the characteristics of the model and the characteristics of the user who interprets the model [20]. The visual design of the model is a primary factor analysis. Recker has conducted research on how process models can be designed to maximize their understanding [21].

References

[1] Deming WE. Calidad, productividad y competitividad: La salida de la crisis. España: Díaz de Santos S.A. 1989.

[2] van der Aalst WMP. Process Mining. Discovery, Conformance and Enhancement of Business Processes. London New York: Springer, 2011.

[3] Verbeek HMW, Buijs JCAM, van Dongen BF, van der Aalst WMP. XES, XESame, and ProM 6. In: Soffer P, Proper E, editors. CAiSE Forum 2010, LNBIP, volume 72; Berlin, Heidelberg. Springer-Verlag; 2010. pp. 60-72.

[4] XESame. (n.d.). Extraído el 20 de abril de 2014 desde http://www.processmining.org. 2014.

[5] Nitro. (n.d.). Extraído el 24 de abril de 2014 desde http://www.fluxicon.org/. 2014

Buijs JCAM. Mapping data sources to XES in a generic way [online]. Master Thesis. Eindhoven, The Netherlands: Technische Universiteit Eindhoven. 2010. [Accedido 16 marzo 2012]. Disponible desde: http://www.processmining.org/_media/xesame/xesma_thesis_final.pdf

[6] Gunter CW. XES Extensible Event Stream standard definition. S.l.: s.n. 2009.

[7] Yzquierdo R. Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso. Tesis de Doctorado, Universidad de las Ciencias Informáticas, Cuba. 2012.

[8] van der Aalst WMP, Weijters T, Maruster L. Workflow mining: discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering; IEEE. 2004;16(9):1128-1142.

[9] Mans RS, Schonenberg MH, Song M, van der Aalst WMP, Bakker PJM. Application of process mining in healthcare: a case study in a dutch hospital. Eindhoven University of Technology, The Netherlands. Berlin, Heidelberg. Springer; 2009. pp. 425-438.

[10] Mans RS, Schonenberg MH, Song M, van der Aalst, WMP, Bakker PJM. Process mining in healthcare: a case study. Eindhoven University of Technology, Netherlands. 2008.

[11] Orellana A, Sánchez Y. Minería de Procesos en salud. Caso de Estudio: modelado de los procesos del área de Emergencia. Twelfth LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014) "Excellence in Engineering To Enhance a Country's Productivity".PRF#254. 2014.

[12] Webster C. EHR BPM: from process mining to process improvement to process usability. Trabajo presentado en EHR Workflow Inc, 2011.

[13] Maruster L, van der Aalst WMP, Weijters T, van den Bosch A, Daelemans W. Automated discovery of workflow models from hospital data. Trabajo publicado en The 13th Belgium-Netherlands Conference on Artificial Intelligence; 2001. pp. 183-190.

[14] Reijers HA, van der Aalst WMP, Weijters AJMM, van Dongen BF, De Medeiros AKA, Song M, Verbeek HMW. Business process mining: an industrial application. Information Systems. 2007;32(5):713-732.

[15] Hernández PM. Aplicación de técnicas de minería de proceso para el control y mejoramiento del proceso de compras nacionales e internacionales de bienes y servicios para proyectos de investigación de la Pontificia Universidad Javeriana. Tesis de Maestría. Pontificia Universidad Javeriana, Facultad de Ingeniería Industrial, Bogotá, 2012.

[16] van der Heijden THC. Process mining project methodology: developing a general approach to apply process mining in practice [online]. Master of Science in Operations Management and Logistics. Netherlands: TUE. School of Industrial Engineering. 2012. [Accedido 20 octubre 2012]. Disponible desde: http://alexandria.tue.nl/extra2/afstversl/tm/Van_der_Heijden_2012.pdf

[17] Rozinat A, van der Aalst WMP. Conformance testing: measuring the fit and appropriateness of event logs and process models. In: Business Process Management Workshops. Berlin, Heidelberg. Springer; 2012. pp. 163-176.

[18] IEEE TASK FORCE ON PROCESS MINING. Manifiesto sobre Minería de proceso [online]. S.l.: s.n. 2011. Disponible desde: http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf

[19] Mendling J, Strembeck M, Recker J. Factors of process model comprehension: findings from a series of experiments. Decision Support Systems. 2012;53(1):195–206.

[20] Recker J, Reijers HA, van de Wouw SG. Process model comprehension: the effects of cognitive abilities, learning style, and strategy. Commun Assoc Inf Syst. 2014;34(9):199-222.

Address for correspondence

Ing. Arturo Orellana García, email aorellana@uci.cu, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-315

Improving Hospital-Wide Early Resource Allocation through Machine Learning

Daniel Gartner[a,b] and Rema Padman[a]

[a] *The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA*
[b] *TUM School of Management, Technische Universität München, München, Germany*

Abstract

The objective of this paper is to evaluate the extent to which early determination of diagnosis-related groups (DRGs) can be used for better allocation of scarce hospital resources. When elective patients seek admission, the true DRG, currently determined only at discharge, is unknown. We approach the problem of early DRG determination in three stages: (1) test how much a Naïve Bayes classifier can improve classification accuracy as compared to a hospital's current approach; (2) develop a statistical program that makes admission and scheduling decisions based on the patients' clincial pathways and scarce hospital resources; and (3) feed the DRG as classified by the Naïve Bayes classifier and the hospitals' baseline approach into the model (which we evaluate in simulation). Our results reveal that the DRG grouper performs poorly in classifying the DRG correctly before admission while the Naïve Bayes approach substantially improves the classification task. The results from the connection of the classification method with the mathematical program also reveal that resource allocation decisions can be more effective and efficient with the hybrid approach.

Keywords:

Artificial Intelligence; Diagnosis-Related Groups; Hospital Costs; Linear Programming.

Introduction

Several recent studies have highlighted the challenges facing many OECD (Organization for Economic Cooperation and Development) countries with rising pressure on healthcare systems due to increasing demand and expenditures for healthcare [1]. As a countermeasure, cost-containment policies such as diagnosis-related groups (DRGs) were introduced [10] and are evaluated continuously [11]. When hospital resources such as beds, diagnostic devices, human resources and operating room availability become scarce, the coordination of patient care and patient flow becomes a challenge. The problem is even more complicated when detailed information about patients who seek hospital admission is captured in an unstructured form or is likely to be incomplete in early stages of care.

DRG-based reimbursement of patient services

Within DRG systems, patients are classified into groups with homogeneous clinical characteristics and resources required during treatment. After the treatment, patients' health insurance pays a fixed amount for treatment depending on the assigned DRG. Within this context, uncertainties in care delivery information play an important role in the effective planning of healthcare processes. In early stages of care, the patients' DRG is unclear, because of potentially missing information and unstructured documentation of patient information using free-text [6]. This introduces a significant challenge to the hospital's operations management when assigning patients to scarce hospital resources while economic objectives such as contribution margin maximization are pursued [5].

The role of hospital information systems to improve hospital-wide patient scheduling decisions

By structuring patient information that is documented when patients seek admission, hospitals can leverage this information to better predict patients' resource requirements. When information about each patient's DRG, resource requirements, and clinical pathway [12] is made available before admission, a hospital can potentially improve the effectiveness and the efficiency of resource allocation decisions.

In this paper, we investigate whether machine learning methods can increase the classification accuracy of an inpatient's DRG as compared to the current approach of using a DRG grouper. In particular, our focus is to classify the patient's DRG before admission in order to improve admission decisions via novel statistical models.

We analyze inpatient data from one year consisting of more than 16,000 records from a 350-bed hospital. Our results show that, in general, machine learning approaches can substantially increase early DRG classification accuracy, especially for elective patients who contact the hospital before admission. Moreover, we demonstrate that machine learning techniques combined with mathematical programming can lead to improved resource allocation decisions.

The remainder of this paper is structured as follows. In the next section, we present two approaches demonstrating how early DRGs can be classified. The first classification approach can be seen as the current approach used by our study site, while the second one uses a Naïve Bayes classifier. We then develop a statistical model for patient admission and scheduling under several constraints, such as precedence constraints between clinical activities and resource constraints. The results section describes the data that we employed in our analysis and presents results of the classification and resource allocation task followed by a discussion. In the final section, we summarize our results and provide directions for further research.

Methods

In the following, we present two DRG classification approaches and a model for combining and planning patient flow with admission decisions.

Early DRG classification techniques

Current approach using a DRG grouper

To determine a DRG at discharge, typically, a simple flowchart-based method is used. The method is implemented in a commercial software called DRG grouper. Figure 1 illustrates the DRG-grouping.

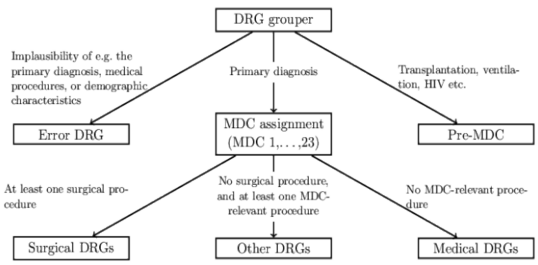


Figure 1 – DRG grouping using a DRG grouper, see Schreyögg et al. [9]

Before the execution of the DRG grouper, parameter values such as the primary and secondary diagnoses, clinical procedures, age, gender, as well as weight in the case of newborns, have to be entered into the software. Diagnoses are coded using the International Statistical Classification of Diseases and Related Health Problems (ICD). The first 3 letters of an ICD code correspond to DRGs. The algorithm first determines one of 23 Major Diagnostic Categories (MDC). These are defined by the primary diagnosis (i.e. the reason for the hospitalization). However, if the primary diagnosis is imprecisely documented, an error DRG will be returned. On the other hand, if the patient has a transplantation, for example, a Pre-MDC (a DRG with high-cost procedures [2]) is returned. After determining the MDC, clinical procedures and co-morbidities lead to the patient's DRG which can be categorized into surgical, medical and other DRGs. Finally, within these categories, the age of the patient, or the weight in the case of newborns, may lead to a different DRG-subtype.

When sufficient information is available, there is valid reason to use a DRG grouper not only at discharge, but at any stage of care in order to obtain real-time information about the patients' DRGs. Therefore, we assume that in every stage of care of a patient and in particular before admission that sufficient information is available to classify the patient's DRG.

Naïve Bayes Classification

Let D be the set of all DRGs, A the set of attributes (e.g. gender, free-text diagnoses) and I be the set of labeled patient instances. The naïve Bayes classifier assumes that all the attributes are conditionally independent given the inpatient's DRG $d \in D$. Under this assumption, the prior probability $p(d)$ of each DRG d is learned from the training data by maximum likelihood estimation, i.e. $p(d)$ is set equal to the proportion of training examples which belong to the class d. Similarly, the conditional likelihood of each value $v_{i,a}$ given instance $i \in I$ and attribute $a \in A$ of each DRG $d \in D$ is learned from the training data by maximum likelihood estimation. This means that $p(v_{i,a} \mid d)$ is set equal to the proportion of training examples of class d which have value $v_{i,a}$ for attribute a in instance i. Afterwards, the classifier assigns the DRG d_i^* to the test instance i by employing Equation (1):

$$d_i^* = \arg\max_{d \in D} \left\{ p(d) \cdot \prod_{a=1}^{|A|} p(v_{i,a} \mid d) \right\} \quad (1)$$

Resource allocation model

In what follows, we present our resource allocation model which is an extension of Gartner and Kolisch's [5] patient flow problem with fixed admission dates. We further extend their model by taking into account admission decisions.

Planning horizon, patients and activities

We have a set of days in which patients can be scheduled and which represents our planning horizon. We have a set of patients who seek admission and can be planned during the planning horizon. The patients have a target admission date and a set of activities to be scheduled within the set of days. For example, an activity can be a computer tomography scan, a surgery or a therapy. Within each patient's set of activities, we have a specific activity which we denote as the discharge activity and which is the final activity of the patient during his/her stay in the hospital.

Time windows, clinical pathways, contribution margin

Each patient's activity has a time window in which it can be scheduled. Once we schedule the discharge activity, the hospital receives a length of stay-dependent contribution margin. The clinical pathway of each patient is represented by precedence relations; each of them can be assigned an integer time lag, which means that the successor activity cannot be scheduled until a certain (waiting) time after the predecessor activity has elapsed. This is necessary to capture recovery times, for example.

Resources, resource capacity and resource requirement

We split hospital resources into a set of day resources and overnight resources as follows: Day resources represent, for example diagnostic devices or the operating room; while overnight resources represent beds since they are typically allocated for at least one night. On the supply-side, each resource has a day-dependent capacity while on the demand-side, each patient's activity has a resource requirement. An activity can require multiple day-resources on the same day, such as when a surgical activity requires the surgery room and the surgeon. Also, an activity does not necessarily require the same amount of time from two distinct resources. For example, a surgical activity may require 100 minutes of surgery room time (including set up time, surgery time, cleaning time, etc.) while the actual time required from the physician may be no more than 80 minutes. Naturally, the documentation task executed by the surgeon follows the surgery and has to be executed before the patient's discharge.

Capacity requirement of overnight resources is typically measured in terms of bed utilization. An overview of all parameters is provided in Table 1 in the Technical Appendix.

Decision variables

Our mathematical model uses binary decision variables to decide whether or not a clinical activity is executed on a day within the activity's time window. Another set of decision variables decides whether the admission of a patient should be avoided. These decision variables inform a decision maker whether or not patients provide the hospital enough of a contribution margin to schedule them along their clinical pathway.

Objective function and constraints

Our mathematical model (its algebraic formulation is shown in the Technical Appendix) maximizes the overall contribution margin of all admitted patients in the hospital by assigning each patient to a discharge date within that patient's discharge time window, see Objective Function (2). Constraints (3) ensure minimum time lags between clinical activities. Note that if we have a time lag of 0 days, the predecessor and successor activity, for example a magnetic resonance tomography and a neurosurgery procedure, respectively, can be performed on the same day. Incorporating the admission variables into the constraints ensures that the constraints are satisfied if the patient is not admitted. Constraints (4) depict the limited capacity of day resources. For each resource and day, capacity demands that all activities performed on that day must not exceed resource capacity. Note that the dimension of resource capacity can be, for example, time or slots in a master surgical schedule or minutes available by a physical therapist. Constraints (5) denote the resource constraints for overnight resources. For each patient, the bed allocation starts with the patient's admission date and the discharge activity releases the bed. Naturally, a patient's resource requirement only sums if he is admitted. Constraints (6) ensure that a patient's activity is scheduled at most one time within the activity's time window. However, once the activity is scheduled, all of the activities corresponding to that patient are scheduled, as required by Constraints (7). Finally, expressions (8)-(9) are the decision variables and their domains.

A bed allocation example

Suppose we have a planning horizon which consists of 7 days and we have two patients to be admitted and scheduled within the planning horizon. Assume that both patients have the actual DRG F21C which means "other surgical procedures with cardiovascular diseases, without complex procedure, without complex skin transplantation and myocutaneous flap surgery of the lower limb admission". The property of the contribution margin function of this DRG reveals that on day 6, it has its maximum. Before that length of stay (LOS), the patient's LOS is below the low LOS trim point. Accordingly, the hospital receives a deductible if the patient is discharged before that day. If the patient's LOS is between the low and high LOS trim point, the contribution margin function is monotonicly decreasing due to LOS-dependent costs for activities such as bed cleaning. We assume that the clinical pathway of the patient requires a minimum time lag of 5 days between surgery and discharge. As a consequence, the time window of the discharge activity has a lower bound for the LOS, which is 5. Since the planning horizon is 7 days, the discharge time window is between 5 and 7 days. We assume that we have one overnight resource and no day resource. The overnight resource is required by both patients. Therefore, only one patient can be admitted and thus, only one admission decision variable is equal to 1. Table 2 in the Technical Appendix gives the algebraic solution of the bed allocation example.

Results and Discussion

In the following, a computational and economic analysis of the scheduling problem is provided.

Data and instance generation

We evaluated our model using data from a mid-sized hospital in the vicinity of Munich, Germany. We joined the hospital data with data from the German institute for reimbursement in hospitals [8]. The latter contains information on the DRG attributes such as low and high length of stay (LOS) trim points, fixed revenue as well as per day reduction and addition. We learned the classifier from January 2011 until June 2011. We ran the simulation experiments from July 1, 2011 until the rest of the year, where each day is solved independently.

Classification results

The classification results of our experiments reveal that the DRG grouper could not classify any of the patients' DRG before admission. One explanation for this phenomenon is that the grouper cannot deal with free-text information. A more detailed analysis revealed that the "error-DRG" 960Z was always assigned to the patients. In contrast, the true positive rate of the Naïve Bayes approach was, on average, 25.3%.

Contribution margin analysis

Our contribution margin analysis revealed that scheduling patients using their DRG, as classified by Naïve Bayes, can substantially increase the hospital's contribution margin, depending on the input parameters of the model and their uncertainty. Our results revealed up to 17% improvement in contribution margin, depending on the planning day. A more detailed analysis reveals that patients classified by the DRG grouper may not be admitted at all because they compete for scarce resources with those patients whose contribution margin is obtained using the Naïve Bayes-classified DRG.

Discussion

The connection between our early DRG classification and the patient admission and scheduling problem should be seen as preliminary results of a sensitivity analysis regarding the extent to which better DRG prediction leads to more effective resource allocation. These models and methods can now be embedded in a real-time decision support system that can inform a decision maker in a hospital about the value of early availability of information about patients to increase classification accuracy and, as a consequence, lead to better resource allocation decisions. This approach for early classification of patients' DRG is not only applicable for elective patients but can also be used for emergency patients, as shown in [6].

Another result of our decision support system based on the mathematical model is that a decision maker can be better informed about patients who would be admitted or scheduled based on a hospital's patient admission and scheduling policy. Furthermore, the decision variables of the model could also reveal that admitting and scheduling patients based on the model's recommendations may lead to higher contribution margins. Identifying these opportunities can also be supported by the approach and tool discussed in this paper which can be included in the platform of Gartner and Padman [7].

Finally, these models and methods provide a rigorous, systematic and evidence-based approach to integrating and analyzing financial objectives of healthcare delivery organizations with operational decisions that are driven by

clinical requirements as well as planning and resource allocation constraints.

Conclusion

In this paper we have presented a novel approach to link classification methods with a discrete optimization model for the problem of scheduling elective patients hospital-wide. The objective is to maximize the contribution margin for elective patients using real-world data from a mid-sized hospital. The results show that if hospitals pursue an objective of making admission decisions based on DRGs, basic machine learning techniques such as Naïve Bayes can lead to more efficient resource allocation decisions as compared to the use of a hospital's current approach using a DRG grouper.

Several streams deserve further research. In order to make the model more applicable in hospitals, our goal is to incorporate overtime flexibility into the model and to consider more specific assignments of patients to a variety of resources, for example, human resources such as surgical teams. Following the approach of Gartner and Arnolds [3], we will also incorporate uncertain clinical pathways into the scheduling process. Preliminary results on a simplified model are shown in prior work [4]. Furthermore, we will evaluate other machine learning techniques such as decision trees, Bayesian networks and logistic regression in combination with the early DRG classification and resource allocation task. Another task is to break down the results into a length of stay analysis, and an evaluation of the confusion matrix in combination with admission and scheduling decisions.

Technical Appendix

Sets, indices, parameters and decision variables

Table 1 provides an overview of the sets, indices, parameters and decision variables.

Table 1 – Sets, indices, parameters and decision variables

Set	Description
A	Activities, Attributes
A_p	Activities for patient $p \in P$
D	DRGs
E_p	Precedence relations
P	Patients
R	Resources
R^d	Day resources
R^n	Overnight resources
T	Days
W_i	Time window for activity $i \in A$
Index/parameter	**Description**
α_p	Admission date of patient $p \in P$
$a \in A$	Attribute
b_p	Specialty requirement of patient $p \in P$
d	DRG
d_i^*	DRG to which instance i is assigned
$d_{i,j}^{\min} \geq 0$	Minimum time lag corresponding to precedence relation $(i,j) \in E_p$
E_i	Earliest day to schedule activity $i \in A$
i	Instance
$i \in A$	Activity
(i,j)	Precedence relation between activity $i \in A$ and $j \in A : i \neq j$
L_i	Latest day to schedule activity $i \in A$
$p \in P$	Patient
$p(d)$	Prior probability of DRG d
$p(v_{i,a} \mid d)$	Conditional probability of DRG d given instance i's value of attribute a
$\pi_{p,t}$	Contribution margin for patient $p \in P$ discharged at day $t \in W_{\phi_p}$
ϕ_p	Discharge activity for patient $p \in P$
$r_{i,k}$	Resource requirement of activity i from resource k
$t \in T$	Day
$v_{i,a}$	Value of instance i's attribute a
Decision variable	**Description**
$x_{i,t}$	1 if activity i is scheduled for day t, 0 otherwise
z_p	1 if patient p is admitted, 0 otherwise

Mathematical program

$$Maximize \sum_{p \in P} \sum_{t \in W_{\phi_p}} \pi_{p,t-\alpha_p} \cdot x_{\phi_p,t} \quad (2)$$

Subject to

$$\sum_{t \in W_j} t \cdot x_{j,t} \geq \sum_{t \in W_i} t \cdot x_{i,t} + d_{i,j}^{\min} \cdot z_p \quad \forall p \in P, (i,j) \in E_p \quad (3)$$

$$\sum_{i \in A : t \in W_i} r_{i,k} \cdot x_{i,t} \leq R_{k,t} \quad \forall k \in R^d, t \in T \quad (4)$$

$$\sum_{p \in P : b_p = k, t \geq \alpha_p} \left(z_p - \sum_{\tau = E_{\phi_p}}^{\min\{t, L_{\phi_p}\}} x_{\phi_p,\tau} \right) \leq R_{k,t} \quad \forall k \in R^n, t \in T \quad (5)$$

$$\sum_{t \in W_i} x_{i,t} \leq 1 \quad \forall i \in A \quad (6)$$

$$\sum_{i \in A_p, t \in W_i} x_{i,t} = |A_p| \cdot z_p \quad \forall p \in P \quad (7)$$

$$x_{i,t} \in \{0,1\} \quad \forall i \in A, t \in W_i \quad (8)$$

$$z_p \in \{0,1\} \quad \forall p \in P \quad (9)$$

A bed allocation example

An example with two patients is presented in Table 1.

Table 2 – A bed allocation example for two patients

t	1	2	3	4	5	6	7
$x_{\phi_1,t}$	-	-	-	-	0	1	0
$\sum_{b_1=1, t \geq 1} \left(z_1 - \sum_{\tau = E_{\phi_1}}^{\min\{t, L_{\phi_1}\}} x_{\phi_1,\tau} \right)$	1	1	1	1	1	1	1
$x_{\phi_2,t}$	-	-	-	-	0	0	0

$$\sum_{b_2=1,l\geq 1}\left(z_1-\sum_{\tau=L_{\phi_2}}^{\min\{\tau,L_{\phi_2}\}}x_{\phi_2,\tau}\right) \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$$

In this example, we assume that we have a planning horizon of $T=\{1,2,...,7\}$ days and that both patients $p=1$ and $p=2$ have an admission date $\alpha_1=\alpha_2=1$ and a discharge time window of $W_{\phi_1}=W_{\phi_2}=\{5,6,7\}$. Moreover, we assume that we have one overnight resource $R^n=\{1\}$ and, for simplicity, no day resource $R^d=\{\}$. The overnight resource is required by both patients. Accordingly, the bed requirements for both patients come up to $b_1=b_2=1$. Assume, that patient $p=2$ cannot be admitted and therefore decision variable $z_2=0$. The table reveals that a bed is allocated for patient $p=1$ from his admission in period $t=1$ until his discharge in period $t=6$. In contrast, patient $p=2$ does not allocate a bed since he is not admitted and, because of constraints (7) he is not discharged.

This study is exempt from IRB because it analyzes de-identified data.

Acknowledgements

The authors sincerely thank the anonymous referees for their careful review and excellent suggestions for improvement of this paper.

References

[1] Organisation of economic co-operations and development, http://stats.oecd.org/Index.aspx?DataSetCode=SHA (last accessed: December 20th, 2014).

[2] Busse R, Geissler A, Quentin W, and Wiley M. Diagnosis-related groups in Europe. McGraw-Hill, Berkshire, 2011.

[3] Gartner D and Arnolds IV. Learning clinical pathways for downstream patient flow planning. Proceedings of 2014 INFORMS Workshop on Data Mining and Analytics (DMA 2014), 2014.

[4] Gartner D, Arnolds I and Nickel S. Improving hospital-wide patient scheduling decisions by clinical pathway mining. MEDINFO 2015 – 15th World Congress on Health and Biomedical Informatics, 2015.

[5] Gartner D and Kolisch R. Scheduling the hospital-wide flow of elective patients. Eur J Oper Res 2014: 223(3): 689-699.

[6] Gartner D, Kolisch R, Neill DB, and Padman R. Machine learning approaches for Early DRG Classification and Resource Allocation. INFORMS J on Computing: Accepted for publication, 2015.

[7] Gartner D and Padman R. E-HOSPITAL – An Environment for Hospital Operations and Services Planning using Information Technology and Algebraic Languages. MEDINFO 2015 – 15th World Congress on Health and Biomedical Informatics, 2015.

[8] Schreyögg J, Stargardt T, Tiemann O, and Busse R. Methods to determine reimbursement rates for diagnosis related groups (DRG): A comparison of nine European countries. Health Care Manag Sci 2006: 9(3): 215-223.

[9] Schreyögg J, Tiemann O, and Busse R. Cost accounting to determine prices: How well do prices reflect costs in the German DRG-system? Health Care Manag Sci 2006: 9(3) 269-279.

[10] Stabile M, Thomson S, Allin S, Boyle S, Busse R, Chevreul K, Marchildon G, and Mossialos E. Health care cost containment strategies used in four other high-income countries hold lessons for the United States. Health Affairs 2013: 32(4): 643-652.

[11] Anderson GF, Willink A, and Osborn R, Reevaluating "Made in America" — Two Cost-Containment Ideas from Abroad. N Engl J Med, 2013: 368(24): 2247-2249.

[12] van de Klundert J, Gorissen P, and Zeemering S. Measuring clinical pathway adherence. J Biomed Inf, 2010: 43(6): 861-872.

Address for correspondence

Daniel Gartner
The H. John Heinz III College
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
dgartner@andrew.cmu.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-320

An eHealth Approach to Reporting Allergic Reactions to Food and Closing the Knowledge Gap

Christopher Munro[a], Aida Semic-Jusufagic[b], Katarzyna Pyrz[c], Philip Couch[a], Audrey Dunn-Galvin[c], Niels Peek[a], Marina Themis[b], Clare Mills[b,d], Iain Buchan[a], Jonathan Hourihane[c], Angela Simpson[b]

[a] *Centre for Health Informatics, Institute of Population Health, University of Manchester, UK*
[b] *Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester and University Hospital of South Manchester, UK*
[c] *University College Cork, Ireland*
[d] *University of Manchester, Institute of Biotechnology, UK*

Abstract

There is an important knowledge gap in food allergy management in understanding the factors that determine allergic reactions to food, in gathering objective reports of reactions in real time, and in accessing patients' reaction-histories during consultations. We investigate how eHealth methods can close this knowledge gap. We report experiences with an online tool for reporting allergic reactions that we have developed as a web application. This application has been successfully validated by participants from Ireland and the UK, and is currently being used in a pilot where participants report allergic reactions in near-real time.

Keywords:

E-Epidemiology; Recall Bias; Web-based Reporting; Allergic Reaction; Food Allergy; Personal Health Record; Self-care.

Introduction

Allergic disease is a growing health risk [1],[2], while its management by clinicians and patients is challenging [3],[4],[5].

Food allergy has reached epidemic proportions in developed parts of the world [6],[7] with up to 20 million European citizens suffering from food allergy [8] and reports of increasing prevalence in developing countries [9]. The reasons for such an increase are not well understood. The estimated worldwide prevalence of food allergy varies according to age with 3-8% reported prevalence among children and 1-3% among the adults [10]. Unsurprisingly, food allergy is a leading cause of anaphylaxis seen in emergency departments across the USA and UK [11],[12].

We can identify the following problems in food allergy:

1. There are gaps in current scientific and societal knowledge, in terms of linking food allergy complaints from patients to evidence. Correlates of risk perception, risk taking behaviours, and psychosocial aspects have been largely overlooked in the literature of the topic.
2. During typical clinician consultations with food allergy patients, the clinician has only a short time (e.g., 20 minutes), and normally no access to consistent detailed information about the patient's previous reactions. Inexperienced doctors may misinterpret both mild and severe allergic reactions. Also, patients describing previous reactions tend to exhibit a recall bias, recalling better the more severe reactions, and failing to report the mild reactions.
3. Food manufacturers often don't have post-marketing information on the incidence of accidental allergic reactions due, for instance, to cross contamination as there is no system in place for reporting real-time allergic reactions in the community setting.

Scientific aim and objectives

Our aim was to improve the capture of objective information on accidental allergic reactions to food in order to address the knowledge gap in food allergy.

Our primary objective was to develop a system that can be used by food allergy sufferers to report information about suspected allergic reactions. This system is intended for collecting data in validation and pilot studies as part of the iFAAM EU FP7 project (integrated approaches to food allergen and allergy risk management), seeking the factors related to food reactions. Data will be integrated into the iFAAM informatics platform Allerg-e-Lab.

Our secondary objective was for this system to be developed as a prototype reporting tool that could be used by clinicians and patients to report and view previous allergic reactions in a clinical context.

Relationship to similar existing systems

There are a number of spontaneous reporting systems for reporting adverse events in areas of blood transfusion [13], medical equipment, drugs or tissues [14],[15], and foods [16]. These tend to be nationwide systems for the capture of reports of adverse reactions for regulatory reasons where there may be a mandatory requirement to enable adverse events to be reported; they never have a clinical function. In contrast, we are developing a system with eventual application in clinical consultations, giving clinicians access to patient-reported historical allergic reactions.

Our approach has been to understand food allergy and the context of reactions as opposed to just a context of foodstuffs (i.e. a register of products leading to adverse events). We capture a range of factors specific to food allergy, photos, food sample descriptions, and a questionnaire regarding the reaction.

Methods

Developing the system requirements

The stakeholders/users of the system are:

- clinicians, researchers, nurses, and study staff at University Hospital South Manchester (UHSM), UK;
- research staff and clinicians at University College Cork, Ireland (UCC);
- study participants recruited online, through the Anaphylaxis Ireland website, the Anaphylaxis campaign website, and UHSM allergy clinic.

The system requirements were developed in an iterative cycle of discussing requirements, implementing a draft solution, and incorporating feedback from stakeholders into refinements.

Questionnaire structure

In order to capture necessary information about food allergy reports from participants, to address our primary objective, the allergic reaction in the community (AlleRiC) questionnaire instrument was developed. This took place between March-October 2013 following 16 focus groups of adults and children diagnosed with food allergies. The instrument was developed in Ireland and the United Kingdom by experienced clinicians and researchers with an active contribution of patients' organisations. Transcripts were then analysed using the grounded theory approach, which involved coding and categorising codes into emergent themes. Items selected from the themes found in the transcripts were taken verbatim and then rephrased into a question format. Items were constructed to reflect the most prevalent characteristics and circumstances of food-allergic incidents as experienced in the real world.

Three experts in the field were consulted in person and via an exchange of e-mails; all disagreements were dealt with by negotiations till consensus was reached. Finally, the recent relevant literature in the field was reviewed to triangulate the above contributions.

In total there are 81 question items including conditional questions. When the participant first logs in they are asked to complete four enrollment questions. The incident questionnaire is then answered for each future reported reaction.

These questions were based on what consumers themselves told us were their main concerns related to the aims of the study, so are grounded in everyday lives and behaviours and therefore one aspect of good construct validity.

The questionnaire captures information about the following factors: subjective intensity measure (Figure 1), location and social context, food/meal/allergen details, physical and psychological symptoms, co-factors such as exercise, trip abroad, period, medication used, and follow-up, labelling and other allergen information, community and professional support, and additional comments.

Validation study

We carried out an initial feasibility study in the form of a validation study, to check that participants are able to use the system to report reactions. A separate group of adult patients diagnosed with food allergies in Ireland and the UK was recruited between December 2013 - February 2014 to evaluate the system by reporting an historic reaction. Participants were recruited online through charities that support food allergy sufferers: the Anaphylaxis Ireland website (www.anaphylaxisireland.ie), the Anaphylaxis Campaign website (www.anaphylaxis.org.uk), and UHSM allergy clinic.

Individual question items of the tool were psychometrically assessed via a novel Evaluative Scale (ES). ES was developed by researchers and patients to allow for the evaluation of five distinct aspects of the prototype questions: ease of understanding, reasonableness, usefulness, ease of answering, and clarity of instructions.

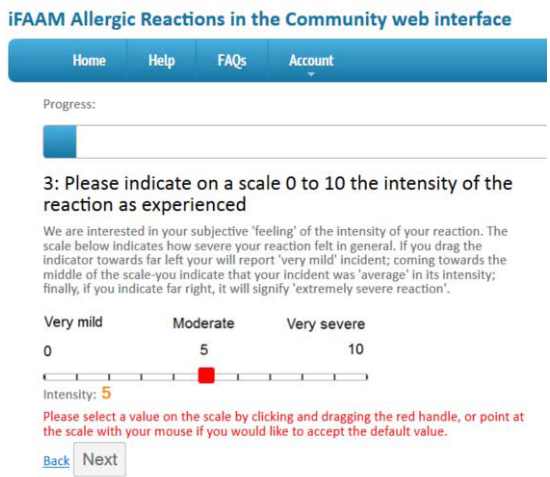


Figure 1 – Screenshot with reaction intensity question

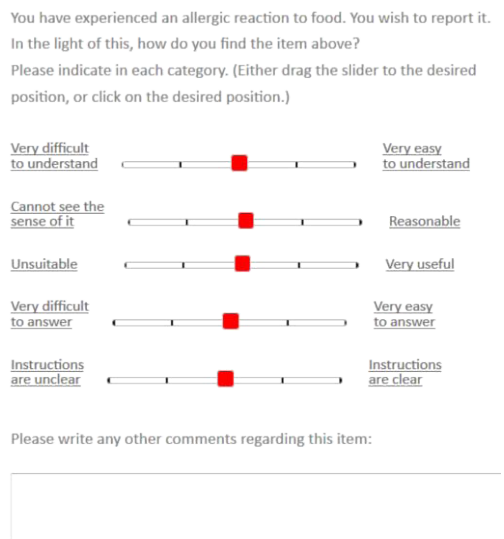


Figure 2 - Evaluative scale as implemented online

Figure 2 shows the evaluative scale as implemented in the AlleRiC system graphically.

ES uses a Likert-type response format on a five-point left-to-right scale range from "0" standing for "very negative evaluation" to "4" standing for "very positive evaluation".

Hence, each question item consists of the question, and the validation section of five drag and drop scales, and a textbox for comments on the question item.

Results

Requirements

Study participant requirements

Study participants need to use the system for three main tasks: 1) when using the system for the first time they should

complete the consent form if they are have not done so in a clinic, view training materials and complete the enrollment questionnaire; 2) if they experience a suspected allergic reaction to food, they will need to complete an incident questionnaire to describe their reaction, if available they can upload photos of skin symptoms or food; and 3) users should be able to access a list of their reported reactions by date, but not extract any data from the system in the current stage of the project.

Clinician requirements

For the pilot study, clinicians needed to use the system to report severity scoring of any skin symptoms. They will view photographs of skin symptoms using a separate system, and select the severity of symptoms in this system using drop-down menus within this application.

Research study staff requirements

Research study staff should be able to carry out activities such as adding and managing participant accounts, and viewing participants responses in the system.

During the pilot stage of the study, the participants are asked to obtain food samples, and upload photographs of packaging where possible. Based on this information, the LanguaL [17] food thesaurus will be used by researchers to describe the food(s) identified by the participant, using relevant faceted classifications. After describing the food(s) with the LanguaL food product indexer [17], the classifications can be exported as XML and the system should allow these to be imported, with the data linked to the reported incident.

The system should allow email notifications to be sent to research staff to assist in organizing the study. These should include: notifying them of participants starting/finishing reporting an allergic reaction; uploading photographs; availability of food samples, LanguaL™ food descriptions being attached to an incident; and clinicians inputting severity scoring based on photos of skin symptoms.

Lead research staff requirements

The requirements of the lead research staff summarize the high level goals for the study using the system.

Properties

Realizing challenging requirements

Here we discuss the realization of requirements into properties of the system.

Recruitment and informed consent

The system allows participants to be allocated to a centre (UK or Ireland for the validation and pilot studies) and will automatically generate their study ID. The system presents an online consent form to participants recruited remotely and not in person through attendance at a clinic.

Workflows

When participants log in for the first time, if they are recruited remotely they are presented with the online informed consent form. The data captured by the system is divided into an enrollment questionnaire and an incident questionnaire. The enrollment questionnaire is done once (gender, age, diagnosed food allergy and asthma diagnosis). The incident questionnaire is completed each time the participant reports a suspected allergic reaction. Once the participant has answered the first question of the incident questionnaire, they can upload photos using the *My reactions* page, which lists the completed/uncompleted allergic reaction reports, with dates.

Once the incident questionnaire is completed, the research study staff can use the system to allocate a food sample to be analysed (if one is available). The LanguaL™ food descriptions and symptom severity scores can be uploaded at this point. The system has a list of incidents, and for each incident the status of questionnaire completion and upload of food descriptions and severity scores can be seen.

Photographs

Participants are requested to upload photos, including foods, packaging, and skin symptoms. The photos of skin symptoms are used by clinicians to carry out a reaction severity scoring. After participants upload the photos they can add descriptive text before confirming the upload. As these photos could be personally identifiable, the system only allows participants to upload the photos, after confirming the photo and providing a description, and displays information that photos have been uploaded but does not allow participants to download or view the photos from within the system. Also, clinicians use a separate secure system to view the photos when they complete the severity scoring.

Customizations for mobile devices and tablets

We made a number of modifications so that participants could access the system using a tablet or mobile device. For example, customizations so that more of the page is available to the question items for tablets. Tablets and mobiles do not have a cursor, so we made changes to the system based on hovering over items so that it would work correctly.

Mobile accessibility/web browser

We considered developing a mobile phone application, however due to resources and the large size of the questionnaire, we decided to develop a web application, with some modifications to enable it to work on mobile devices and tablets. However it requires use of a web browser and a network connection to communicate with the server.

Developing a phone application would have the advantage that participants could complete the questionnaire without a network connection and submit the data once a network connection is found. Participants are reminded not to report reactions until they have fully dealt with their symptoms, and the system is a prototype, so we thought this was an acceptable trade-off. Developing a mobile application would consume more resources as we would have needed to use a number of different technologies to produce software for each operating system. With a web application we only need to develop and support one system.

Participants have a *My reactions* page, listing their reported reaction incidents. Once participants start reporting a reaction and have answered the first question item, an incident is added to this list, and they can click a link to upload photos. If they log into the system from a mobile or tablet device supporting the HTML5 media capture *camera* tag they can upload a photograph directly from their device. As there are at most 42 items, we expect most participants to complete the questionnaire via a web browser at a computer or via a tablet, and use a mobile phone primarily as a camera and not to answer the questionnaire.

Source documentation

There is a requirement from clinical staff to be able to print off source documentation, including online consent forms and report reaction incident questionnaire responses.

As the system does not store any names of participants, when printing the consent form (for remotely recruited participants) the clinician can input the names of the participant, and clinical staff and print a hard copy. The system then records that the consent form has been printed.

Validation study results

There were three separate units of analysis at this stage of validation: 81 items of the prototype evaluated using an ES; ES as a psychometric construct; and the initial questionnaire as a psychometric construct.

Thirty-nine adults from Ireland and UK, diagnosed with food allergy, evaluated the prototype online. Tables 1 and 2 show the demographics of the 39 adults, with Figure 3 showing the age distribution. Fifty-four percent of the adults also had asthma, the most common allergies being peanut and treenut.

Individual items were evaluated positively or very positively by participants (60-70% of positive scores across the ES). Qualitative thematic analyses identified four main topics of concern, however these did not relate to the prototype or any of the items, so this data was not used in the subsequent refinement of the prototype.

Reliability of ES was assessed by Cronbach's Alpha statistic with any figure over 0.7 being regarded as good validity. ES yielded .991 Cronbach's Alpha demonstrating a very good internal reliability.

Acceptance during validation study

Our aim is to develop a system that is used to report reactions, here we assess basic measures of the use of the system: the number of questions answered, approximate time to complete the questionnaire, and the number of users evaluating the system by reporting a reaction. This is linked to what users want/need to be able to report accurately and conveniently.

Sixty-three usernames were allocated for the validation study, 39 questionnaires were successfully completed. For the uncompleted questionnaires, only four people started but failed to complete, the remainder did not start the questionnaire.

We recorded the time the first and last question items were submitted to use the difference to crudely approximate the time taken to complete a response. This assumes that participants complete the questionnaire in one session, without stopping. This measure will not capture the duration correctly if for example they log out of the system and log back in later to resume the questionnaire, or pause completing the questionnaire. We have discounted any reporting times over one hour which we can be reasonably sure involved leaving the questionnaire and returning later. The average time for the remaining completion times was 27.5 minutes (minimum 10, maximum 62 minutes). The time to complete the validation questionnaire is likely to be significantly longer than the pilot study, due to the evaluative scale present for each question item; this scale is not used during the pilot study.

Participants did not need to complete all 81 items due to the conditional nature of some question items, the maximum number of items in a completed response was 42, and minimum 29 items. Tables 1 and 2 show baseline characteristics of the participants.

Table 1– Baseline characteristics during validation study

Gender	Ireland	UK	Total
Male	3	2	5
Female	18	16	24
Total	21	18	

Table 2 — Diagnosed allergies during validation study

Allergy	Participants with allergy	Allergy	Participants with allergy
Peanuts	28%	Milk	7%
Treenuts	24%	Wheat	3%
Seafood	9%	Fruit/veg	13%
Egg	11%	Other	5%

Discussion

We developed a prototype system to enable the reporting of suspected allergic reactions in near-real time, capturing a range of information about the context and nature of the reactions. This system has been successfully used in a validation study with 39 participants from the UK and Ireland using it to report an historical reaction.

Our system has some limitations, for example, as it is accessed via a web browser, we require network connectivity to access the system. Participants are asked to upload photographs if these are available. This can be done either by uploading the photos via a web browser from a computer, or, by logging into the system on a phone/tablet with a web browser, in which case the camera on the device can be used. The results of the pilot study will enable us to see whether participants upload many photos; we did not ask participants to upload photos during the validation study.

The validation data indicated some issues with distribution resulting from the voluntary nature of the recruitment, most notably, the sample was not representative of a wider population with 90% of all participants being female. This gender bias will need to be addressed in the next stages of the validation process.

The study asks participants to deal with their symptoms before reporting a suspected allergic reaction, this will result in a delay between the reaction and reporting the reaction. The amount of time passing before reporting symptoms influences the reported symptoms. Therefore as the system captures the time of reaction as a question, and records the time the questionnaire was submitted, in the analysis we could control for time lag between reaction and report.

In regard to the length of the questionnaire, and the validation process, it is important to note that during validation participants only used the system once. For the pilot study they may report along the full spectrum of severity – this is novel because usually only the severe end of the spectrum is presented in accident and emergency hospital departments. This type of information may be used to inform future integrated dynamic models of 'severity' and behaviours.

The validation and pilot studies have involved recruiting participants through clinics, and anaphylaxis campaign websites, and these populations may not be representative of the general population in terms of engagement with the system.

AlleRiC is currently undergoing further validation in its pilot stage, which involves live-reporting of food-allergic reactions.

This phase runs from October 2014 for a period of 18 months. The recruitment strategy for the UK is the University of Manchester population, the UHSM allergy clinic, and Anaphylaxis campaign, and for Ireland includes University College Cork population (there is no clinical involvement in Ireland), GPs, private allergy clinics, Anaphylaxis Ireland members, and through social media. Both UK and Ireland use online recruitment. The eligibility criterion are: adults aged over 18, capable of giving consent, with a physician diagnosed

food allergy. Participants will use the system to report suspected allergic reactions over this period. The resulting dataset will be analysed quantitatively (aspects of food eaten, treatment etc.). A qualitative analysis will be done on suitable text fragments, e.g., open ended text questions, to investigate participant experiences related to allergic reactions to food.

After the completion and analysis of results of the pilot study, cross cultural validation will be undertaken using a multi lingual version of the system across four distinct geographic regions in Europe.

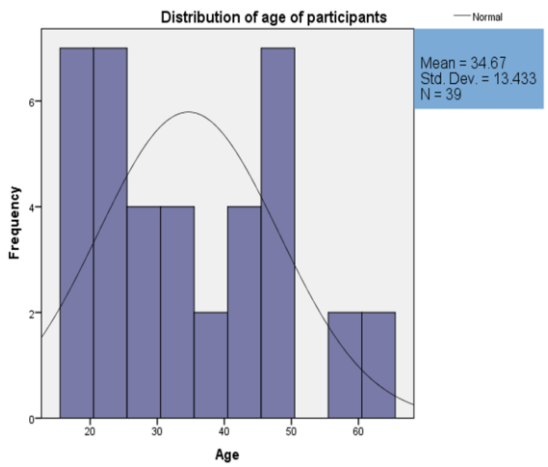


Figure 3 –Age distribution for initial feasibility study

Conclusion

A web based system, AlleRiC has been developed and now enables reporting of suspected allergic reactions to food in near-real time. The system has been used in a validation study with 39 adults from the UK and Ireland who have diagnosed food allergies – successfully reporting historical reactions and evaluating the system. The system is now deployed in a pilot study in the UK and Ireland where participants are reporting reactions prospectively in near real time.

As more people spend more of their lives online there is an epidemiological opportunity to tackle recall bias in ways that might otherwise increase sampling bias due to diferences in technology access. We have used this e-epidemiology tipping-point to address a big gap in food allergy knowledge.

Acknowledgements

The iFAAM project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No KBBE.2012.2.4-04. Supported also by MRC Health eResearch Centre, Farr Institute (MR/K006665/1). We thank Anaphylaxis Ireland, the Anaphylaxis Campaign UK, and the study participants for their help with the studies.

References

[1] Madsen, C.B., Crevel, R.W.R., Mills, C. & Taylor, S.L. (Eds.) (2014). Risk Management for Food Allergy.

[2] Mills, E.N., et al. (2007). The prevalence, cost and basis of food allergy across Europe. Allergy. *62(7),* 717-722

[3] DunnGalvin, A., et al. (2010). Longitudinal validity and responsiveness of the Food Allergy Quality of Life Questionnaire - Parent Form in children 0-12 years following positive and negative food challenges. Clinical and Experimental Allergy. 40(3), 476-485.

[4] Grossman, S. L., et al. (2013). Anaphylaxis knowledge and practice preferences of pediatric emergency medicine physicians: a national survey. Journal of Pediatrics. 163(3), 841-846.

[5] Gupta, R. S., et al. (2010). Food allergy knowledge, attitudes, and beliefs of primary care physicians. Pediatrics. 125(1), 126-132.

[6] Berin MC, Sampson HA. Food allergy: an enigmatic epidemic. Trends in Immunology.

[7] Burks AW, Tang M, Sicherer S, Muraro A, Eigenmann PA, Ebisawa M, et al. ICON: food allergy. J Allergy Clin Immunol 2012; 129:906-20.

[8] Jones SM, Burks AW. The changing CARE for patients with food allergy. J Allergy Clin Immunol 2013; 131:3-11; quiz 2-3.

[9] Hu Y, Chen J, Li H. Comparison of food allergy prevalence among Chinese infants in Chongqing, 2009 versus 1999. Pediatrics International 2010; 52:820-4.

[10] Longo G, Berti I, Burks AW, Krauss B, Barbi E. (2013) IgE-mediated food allergy in children. The Lancet. 382 (9905), 1656-64

[11] Gibbison B, Sheikh A, McShane P, Haddow C, Soar J. Anaphylaxis admissions to UK critical care units between 2005 and 2009. Anaesthesia 2012; 67:833-9.

[12] Ross MP, Ferguson M, Street D, Klontz K, Schroeder T, Luccioli S. Analysis of food-allergic and anaphylactic events in the National Electronic Injury Surveillance System. Journal of Allergy and Clinical Immunology 2008; 121:166-71.

[13] Medical and Healthcare Products Regulartory Agency UK, Serious Adverse Blood Reactions and Events (SABRE) http://www.mhra.gov.uk/Safetyinformation/Reportingsafetyproblems/Blood/

[14] Health Canada. MedEffect™. http://www.hc-sc.gc.ca/dhp-mps/medeff/index-eng.php

[15] United States Food and Drug Administration. MedWatch online voluntary reporting form. http://www.fda.gov/Safety/MedWatch/

[16] Løvik, M., Namork, Fæste, C., and Egaas, E. (2004) The Norwegian national reporting system and register of severe allergic reactions to food. Norsk Epidemiologi. 14 (2), 155-160.

[17] LanguaL™ International framework for food description. www.langual.org

Address for correspondence

Christopher Munro, PhD, MSc
Centre for Health Informatics, The University of Manchester
christopher.munro@manchester.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-325

PancreApp: An Innovative Approach to Computational Individualization of Nutritional Therapy in Chronic Gastrointestinal Disorders

Konrad Stawiski[a], Alicja Strzałka[a], Anna Puła[a], Krzysztof Bijakowski[b]

[a] *Faculty of Medicine, Medical University of Lodz, Lodz, Poland*
[b] *Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, Cracow, Poland*

Abstract

Medical nutrition therapy has a pivotal role in the management of chronic gastrointestinal disorders, like chronic pancreatitis, inflammatory bowel diseases (Leśniowski-Crohn's disease and ulcerative colitis) or irritable bowel syndrome. The aim of this study is to develop, deploy and evaluate an interactive application for Windows and Android operating systems, which could serve as a digital diet diary and as an analysis and a prediction tool both for the patient and the doctor. The software is gathering details about patients' diet and associated fettle in order to estimate fettle change after future meals, specifically for an individual patient. In this paper we have described the process of idea development and application design, feasibility assessment using a phone survey, a preliminary evaluation on 6 healthy individuals and early results of a clinical trial, which is still an ongoing study. Results suggest that applied approximative approach (Shepard's method of 6-dimensional metric interpolation) has a potential to predict the fettle accurately; as shown in leave-one-out cross-validation (LOOCV).

Keywords:

Personalized medicine; nutritional therapy: machine learning; chronic gastrointestinal disorders.

Introduction

The experience of pain and complaints associated with gastrointestinal disorders determines patient's fettle and has an impact on the recovery rate. Nutrition therapy, although complex in nature, has a meaningful role in various gastrointestinal disorders such as chronic pancreatitis [1], inflammatory bowel diseases (Leśniowski-Crohn's disease and ulcerative colitis) or irritable bowel syndrome.

Pancreatitis is commonly defined as a continuous, chronic, inflammatory process of the pancreas, characterized by irreversible morphologic changes. For the majority of patients with chronic pancreatitis, abdominal pain is the most common symptom. This intermittent pain may worsen after eating or drinking. There is a sizeable list of reported etiological factors of pancreatitis which can be categorized into: (1) toxic and metabolic (connected mostly with alcohol intake), (2) genetic, (3) autoimmune, (4) recurrent acute or severe acute pancreatitis, (5) obstructive and, lastly, (6) idiopathic. According to a paper from 2008 by Shallu et al. the main issues in chronic pancreatitis are exocrine insufficiency and pain – which are also the two main reasons of diet regime and malnutrition in pancreatitis patients. Unquestionably, the dietary regime can soothe inconveniences and thus self-control is as crucial as clarity of recommendations [2]. However, guidelines may not be appropriate for every individual, thus personalization is needed. Lack of the individualization can affect directly the compliance, as guidelines are usually complex and can be misinterpreted by patients. Personalized therapy applied in other scenarios, has recently shown positive results [3, 4, 5].

Individualized diet can increase fettle and improve the compliance of patients suffering from chronic gastrointestinal disorders. Diet-related patterns are also being constantly searched for in inflammatory bowel diseases and in the irritable bowel syndrome. This is driven by the fact that etiology of these diseases is still unclear.

The aim of this study is to develop and evaluate a solution with the purpose of supporting patients' self-control and adjust diet to the best possible fettle based on an individualized machine learning model. The proposed solution iscalled – PancreApp (fig. 1).

Figure 1 – Logo of PancreApp project.

Materials and Methods

This paper describes the process of idea development and application design, feasibility assessment using a phone survey, preliminary evaluation on 6 healthy individuals and the earliestresults of the clinical trial, which is still an ongoing study (21 patients involved so far).

Idea development and application design

PancreApp is designed as a multiplatform, cloud-based digital diet diary, which not only enables regular control of the disease, but also tries to predict diet-related changes of fettle using computational methods. The application considers the details of diet and associated disease manifestations as input. The aim of the data processing is to provide patients with an individually computed multidimensional profile of their

disease which can be applied in the estimation of fettle for the next meals. In this model continuous acquisition of new data enables constant calibration of prediction; in general a concept of machine learning algorithms. (fig. 2)

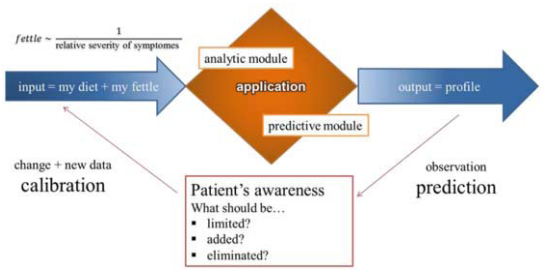


Figure 2 – Graphical representation of PancreApp core concept.

We developed two distributions of the software. The first one, for Microsoft Windows OS (using C# programming language based on .NET Framework CRL, Microsoft Visual Studio 2013 Professional) and the second one - a mobile version for Android OS users (JAVA programing language). Applications, being in bilateral connection with server (python-based, HTTPS protocol, XML format), allow the user to use software in every scenario. The application is harvesting data as the user is determining the diet and fettle after the meal (on a scale).

Entering data is facilitated by the fact that food products available in the market are almost always described quantitatively, based on basic ingredient content. It offers a possibility to count how much of the basic compositions (fat, carbohydrates, fiber, proteins and energy) the users consume during a selected meal and also provides patients with a possibility to archive their diet faster (via simple selection, rather than manual calculation). Our product database contains over 2200 products that are available in the Polish market. When the diet is known, users can define their fettle (associated with a particular meal). The fettle can be determined by the user on a scale from 1 to 10 (where 10 is the best) and should reflect the severity of symptoms. Screenshots of the general user interface are shown in figure 3.

The application, in addition to the prediction model, contains a reach analysis module wherein the patient can find statistics and charts visualizing his diet. The user can ,for example, put together the amount of of consumed fat and the associated fettle. This particular relationship may be clinically relevant in chronic pancreatitis.

The prediction model allows the user to estimate the fettle of future meals based on previous meals and fettle (training set). Its screenshot is presented in figure 4. Among all the methods, we were looking for a simple one that would give users results quickly, even on devices where central processing units have low clock rates. -, PancreApp's prediction model is using Shepard's method of 6-dimensional (multivariate) metric interpolation; as inverse distance weighting is the simplest interpolation method. Specifically, multi-dimensional euclidean distance is calculated and then applied for weighting. Fettle is predicted on the previously described scale and its value is specific for each patient.

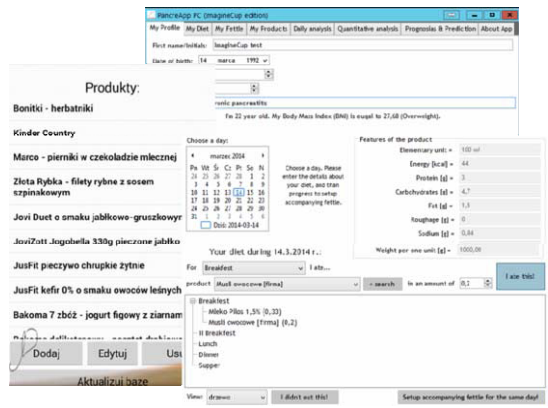


Figure 3 – Screenshots of PancreApp's general user interface. Right-sided pictures present the application for Microsoft Windows OS, left – for Android OS.

The latest version of the application is available for MEDINFO participants to download and test. Links are attached in the references. [6]

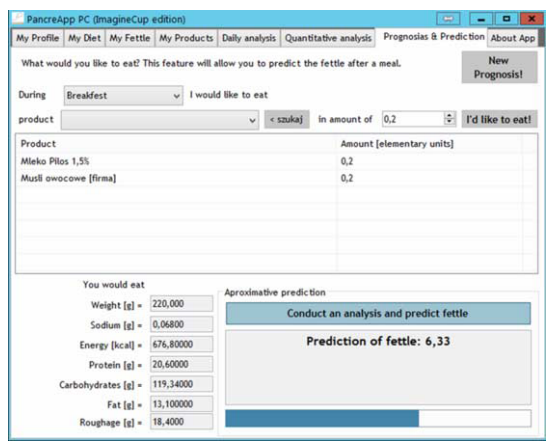


Figure 4 – Example of prediction.

Preliminary evaluation

Preliminary evaluation has been conducted on 6 healthy individuals that were using the software on a daily basis for 2 weeks. We focused on a comparison of relative difference between a real fettle set by the user and a potential prediction that could be calculated using the rest of the training set. This applied technique is henceforth referred to as the leave-one-out cross-validation (LOOCV). STATISTICA (StatSoft Inc.) and our proprietary software were used for data analysis.

We have also conducted a phone survey on 9 patients with chronic pancreatitis after pancreatic drainage/resection during the last 3 years to assess if the proposed idea reflects the needs of the target group. We asked about pain, depression (PHQ-9), food intolerance and access to a personal computer or to a mobile phone with Android.

Clinical trial

Since October 2014 this project has been financed by the Polish Ministry of Science and Education as a grant for student researchers. We enrolled adult patient-volunteers with chronic pancreatitis, inflammatory bowel diseases

(Leśniowski-Crohn's disease and ulcerative colitis) or irritable bowel syndrome. The only requirement was possessing a device with continuous access to the internet and the completion of the application tutorial. Enrolled patients were randomly divided into the control and experimental group and were then using the application for the 3 following weeks. The experimental group was able to use the prediction model, while the control group could use PancreApp only as a simple diet diary. At the beginning and at the end of the trial we measured patients' basic physical parameters and PHQ-9 score (assessment of depression that may affect compliance). To this day, 21 out of 100 patients (planned) have been enrolled into the trial, while 10 of them have finished it. In the analysis we want to assess how well this algorithm can predict fettle (by correlation coefficient and root mean squared error as well as LOOCV), how often the prediction model is used (frequency) and if patients experience time-related changes in fettle. We also want to analyze z-scores of fettle, compared to the amount of consumed basic compositions.

Results

During preliminary evaluation, the study group consisted mainly of young people with normal BMI (mean: 20.06; mean age; 24.3; mean energy consumed: 1408.4 kcal). Related-Samples Wilcoxon Signed Rank test showed that median difference between the well-being (fettle) value described by the tester and the well-being value predicted using Shepard's algorithm (based on all data except for tested; LOOCV) was not statistically significant (p=1.00). The absolute error of prediction was equal to 0.6454, while the relative error of prediction equaled 7.59% of fettle value.

The survey showed that studied patients with pancreatitis (mean age: 52.4) have a depression according to PHQ-9 score (fractions: minimal 0.11 vs. mild 0.55 vs. moderate 0.33). Food-specific intolerance was noted by 8 out of 9 surveyed patients. All surveyees were interested in participation. Only one of surveyed patients had a phone with Android OS. The rest of the patients would prefer using PC with Windows OS.

In this paper we would also like to describe the early results of the clinical trial that will end in May 2015. As described above, 10 patients finished the trial so far, however only 4 of them completed it successfully (used the application over the full testing period). The analysis of patients who dropped out has shown that 2 out of 6 patients were using the application irregularly (despite of reminders), 3 out of 6 never used the application after training meetings and were unreachable for further contact, while one patient was affected by a compatibility issue between .NET Framework 4.5, ClickOnce installer and Windows XP. None of the enrolled patients have , so far, decided, to use the Android version of the application. There was no significant difference between patients who completed the study and patients who dropped out, also in the PHQ-9 score. Patients presented a mean age of 29.8, weight of 68.6 kg, height of 176.2 cm and PHQ-9 score of 8.

The analysis of the 4 patients who have completed the trial so far, has shown good properties of Shepard's method. Two of those patients have Crohn's disease, one has ulcerative colitis and one irritable bowel syndrome. The chart in figure 5 presents the comparison between the predicted value of fettle and the real one, set by the user. Please note the distance between native data and ideal prediction showed by the red line. It is visible that a 95% confidence interval of regression almost contains the line of ideal prediction. The root mean squared error is equal to 0.80, while Spearman's correlation coefficient between predicted and real values is equal to 0.87 (p=0.000). Time-related changes and the usage of the prediction model will be assessed when the trial is completed. The performed analysis considered only whole days, without division into separate mealtimes.

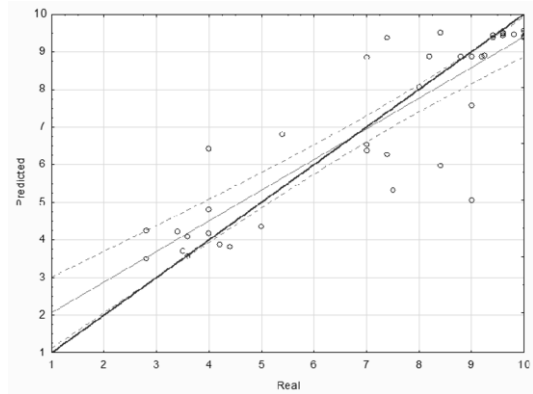


Figure 5 - Comparison between the predicted value of fettle and the real one. The solid thick line presents ideal prediction (f(x) = x), while the thin one (with concomitant dashed lines) shows the linear regression (and 95% confidence intervals). Analyses were performed for whole days. Please note that days without determined fettle change were excluded.

Discussion

Doctors of medicine frequently complain about therapeutic compliance. Lack of faith in treatment is one of the factors predisposing reduced compliance [7], while the therapeutic effect may be strictly dependent on feedback from a patient. PancreApp targets both problems. Machine learning and data-mining techniques are very promising and widely used in various aspects of science. The algorithm applied here is one of the simplest. It can, however, be quickly deployed in an everyday scenario (on low-end devices). It is also necessary to notice the need to test and compare different regression methods on data harvested from this study (retrospectively).

So far, the idea of PancreApp is appreciated by patients and doctors. Despite the fact that we have analyzed early results, from a clinical point of view the prediction accuracy of PancreApp is good. However, several limitations should be noted. Firstly, usage of fettle as a self-assessment of clinical state is disputable, and modern medicine is going to be focused on the quality of life rather than on simple survival. Also, the Hawthorne effect, or selection and technology bias cannot be excluded as a convenient sample is being enrolled. Patients without basic computer skills cannot either benefit from this study. However, based on the results, PancreApp seems to be intuitive enough for usage. Current compliance is in the expected range, as all patients (qualified so far) are in a remission state.

It is also notable that not one patient has chosen an Android device so far (even if they have such an opportunity), suggesting that personal computers are more approachable when it comes to complex data entry.

Conclusion

In this paper we present an innovative approach to computational individualization of nutritional guidelines. Early results from evaluation have shown that the applied approach has a potential to predict the fettle accurately and with low error. A possible broad spectrum of application (pancreatitis, obesity, kidney and liver disorders) encourages us not only to finish the current trail, but also to check whether PancreApp could be applied in different scenarios (diseases). Furthermore, PancreApp can be used in evidence based medicine as an easy available data-gathering service. The main advantage is the flexibility achieved by individualized models. Data collected by this tool could also be used in retrospective analysis of studied disorders, and as a source of valuable inferences. The software has been approved by the local ethical committee. More details about the project can be found at www.pancreapp.pl

Acknowledgments

- Dr. Piotr Hogendorf, M.D., Ph.D. – for being our clinical tutor;
- Polish Ministry of Science and Education – for faith in medical students, trust and financing;

References

[1] Balliet WE, Edwards-Hampton S, Borckardt JJ, Morgan K, Adams D, Owczarski S, Madan A, Galloway SK, Serber ER, Malcolm R. Depressive Symptoms, Pain, and Quality of Life among Patients with Nonalcohol-Related Chronic Pancreatitis. Pain Res Treat. 2012;2012:978646.

[2] Grant JP. Nutritional support in acute and chronic pancreatitis. Surg Clin North Am. 2011 Aug;91(4):805-20, viii.

[3] Lasierra N, Alesanco A, Guillén S, García J. A three stage ontology-driven solution to provide personalized care to chronic patients at home. J Biomed Inform. 2013 Jun;46(3):516-29.

[4] van der Heijden M, Lucas PJ, Lijnse B, Heijdra YF, Schermer TR. An autonomous mobile system for the management of COPD. J Biomed Inform. 2013 Jun;46(3):458-69.

[5] Spat S, Höll B, Petritsch G, Schaupp L, Beck P, Pieber TR. Automatic system testing of a decision support system for insulin dosing using Google Android. Stud Health Technol Inform. 2013;186:187-91.

[6] Windows installer: http://link.konsta.com.pl/p0 and products database: http://link.konsta.com.pl/1n; Android installer: http://link.konsta.com.pl/p1; please use username: medinfo and password: medinfo

[7] Jin J, Sklar GE, Min Sen Oh V, Chuen Li S. Factors affecting therapeutic compliance: A review from the patient's perspective. Ther Clin Risk Manag. 2008 Feb;4(1):269-86.

Address for correspondence

Konrad Stawiski, konrad@konsta.com.pl

Alicja Strzałka, alicja.strzalka@stud.umed.lodz.pl

General contact with authors: autorzy@pancreapp.pl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-329

An Intelligent Ecosystem for Providing Support in Prehospital Trauma Care in Cuenca, Ecuador

Cristian Timbi-Sisalima[a,c], Edgar B. Rodas[b,c], Juan C. Salamea[b,c], Hernán Sacoto[b,c], Diana Monje-Ortega[c], Vladimir Robles-Bykbaev[a,c]

[a] *Assistant professor, GI-IATa, Universidad Politécnica Salesiana, Cuenca – Ecuador*
[b] *Assistant professor, Universidad del Azuay, Cuenca – Ecuador*
[c] *Grupo de Investigación en Telemedicina/Telesalud – CEDIA.*

Abstract

According to facts given by the World Health Organization, one in ten deaths worldwide is due to an external cause of injury. In the field of pre-hospital trauma care, adequate and timely treatment in the golden period can impact the survival of a patient.

The aim of this paper is to show the design of a complete ecosystem proposed to support the evaluation and treatment of trauma victims, using standard tools and vocabulary such as OpenEHR, as well as mobile systems and expert systems to support decision-making. Preliminary results of the developed applications are presented, as well as trauma-related data from the city of Cuenca, Ecuador.

Keywords:

Clinical Decision Support Systems; Emergency Medicine; Mobile Applications; Injuries.

Introduction

Trauma is considered a pandemic worldwide that mostly affects young persons. Therefore, it has a negative impact on the production system of any nation [1][2]. This situation is not different in Ecuador, where since 1995 the injury burden has been documented in a national statistics database [3][4][5]. In this database, external causes of injury are not grouped as a single category, but rather are identified as separate etiologies such as traffic accidents, falls, violence, burns, animal bites, among others [6][7].

According to the Ecuadorian Institute of Statistics and Census (INEC), in 2011, the leading cause of death was diabetes mellitus with 7.15%, while land transportation accidents were the fifth leading cause, and homicides occupied seventh place. f we add up all the traumatic events, trauma becomes the leading cause of mortality at 8.76% [8].

For almost two years now, trauma care in the province of Azuay ha-s been provided at the Vicente Corral Moscoso Regional Hospital, by an Integrated Security System (SIS ECU 911) [9], which coordinates relief agencies in pre-hospital care (PHC) such as the Ministry of Public Health (MSP), the Fire Department, and the Ecuadorian Red Cross. During this period, trauma-related emergencies corresponded to 46% of all care provided, and 70% of this was managed at the public health hospital [4].

On those grounds, we have analyzed the interaction between the pre-hospital primary attention services and the transfer to the hospital. Based on this general analysis (response times, false trauma alerts, etc.) we have determined that the communication system between ambulances and hospitals requires improvements. Therefore, implementation of an improved data transmission system was undertaken. The aim of this paper is to describe how an ecosystem-related software can accomplish this improvement in patient care. Likewise, in this paper we will present preliminary results of the first support application developed; a mobile tool to improve the trauma data collection, communication with the hospital, route planning, and arrival estimation.

The rest of the paper is organized as follows: first, we present a general overview of the proposed ecosystem and its main components. Next, we detail the experiment we carried out as well as the preliminary results. Thereafter, we present some relevant research and discuss some applications of information systems for decision support and trauma care. Finally, our conclusions and future work are presented.

Methods

Trauma care is a complex field that currently requires an integrated solution able to face a wide spectrum of needs such as the following:

- Provide effective pre-hospital attention with the aim of reducing mortality.
- Raise awareness in people that are more prone to suffer accidents related with trauma (persons under 45 years).
- Have robust mobile systems to provide primary healthcare services during accidents.
- Have Information and Communication Technology (ICT) -based tools to support the learning process of medical students.
- Use standardized vocabulary with the aim of sharing and exchanging the patient's information.
- Use knowledge-based models for storing the clinical information.

On those grounds, we have designed an ecosystem that relies on a MLST system (mechanism of injury, suspected lesions, vital signs and effected treatment), and is able to manage critical information from pre-hospital care. Likewise, our proposal includes a prototype for the automatic capture of vital signs from equipment present in the ambulance, and an intelligent system with decision-making algorithms for the categorization of severity of trauma and prioritization of trauma alert. The captured information of the patient and his/her medical condition is sent to the nearest hospital as well as the geographical position of the ambulance, and the arrival time.

With the aim of providing complete support for trauma care, bearing in mind the needs mentioned above, we have designed a comprehensive ecosystem named SINATRA (Intelligent System to Support Trauma, for its acronym in Spanish). This model (Figure 1) relies on a layer-based approach that easily allows integrating more modules/services to cover new areas of trauma care, and supports the requirements of doctors, paramedics, students, and patients. The main objective of our ecosystem is to provide support to carry out activities and tasks that require decision support, and services to share, exchange and analyze information related to trauma care.

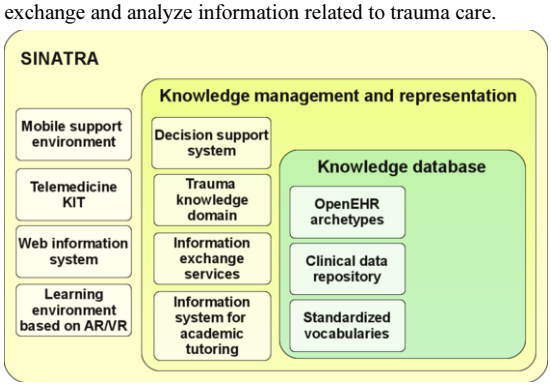


Figure 1– The proposed ecosystem for trauma support.

In this section we will provide a general description of the elements that make up the ecosystem as well as the preliminary achieved results using the mobile application.

Support tools based on ICTs

This layer is based on a set of ICTs tools that provide the interfaces to register medical information, access knowledge databases, and perform training with medical students:

- The *mobile support environment* is a set of mobile applications that provide support in the execution of several tasks such as registering critical information generated in the ambulance (pre-hospital care), remote patient monitoring, and registering and analyzing hospitals infrastructure for trauma care (staff, equipment, protocols, etc.).
- The *telemedicine KIT* is an electronic device that records patient vital signs (respiratory rate, blood pressure, pulse, and oxygen saturation) when he/she is being transported to the hospital.
- In order to manage the information stored in the knowledge database, the ecosystem includes a *web information system*. This tool is used to query a database, register new cases, generate reports, provide statistics, and support some data mining tasks like cluster creation according to several variables (geographical area of accident, demographic patient information, type of injury, etc.). Some components are based on OpenEHR archetypes and templates (www.openehr.org).
- With the aim of reinforcing some learning activities as well as raising the awareness about trauma, the ecosystem includes a *learning environment based on augmented reality and virtual reality (AR/VR)*. This environment can be used during classes at universities, or awareness campaigns for young drivers, seniors, and the general population.

Knowledge management and representation

This layer relies on a set of interfaces and provides several services to manage information. Additionally, in this layer it is possible to incorporate intelligent modules to cover areas related with machine learning, decision support, pattern recognition, data mining, and learning support, among others. Some of the most relevant elements are the following:

- The *decision support system* uses decision trees, neural networks and case-based reasoning to support decision-making in various aspects such as determining the type of alert that is sent to hospital after pre-hospital care (analysis of trauma, patient condition, etc.), allocation of resources for patient care (both human and material), as well as diagnosis and treatment.
- The *trauma knowledge domain* contains a complete set of protocols, procedures and trauma-related ontologies (initial management of trauma, shock, traumatic brain injury, infections, etc.).
- With the aim of providing the necessary tools to access to knowledge bases and share information with applications from other platforms, the ecosystem includes a set of *information exchange services*. These services include the templates generated from archetypes, exporting functionalities, and querying functions, among others.
- For a medical student that is in the academic training phase, it is very important to address real cases related to specific trauma situations. For this, our proposal incorporates an *information system for academic tutoring* that - can - generate random cases of patients with different cases of injuries. With the aim of saving the patient life or provide adequate treatment, the student should make a decision, according to the presented information. The system will rate the decision and will present a valid solution.

Knowledge database

- The *OpenEHR archetypes* allow us to model the different procedures, observations, instructions and data required to register information related to trauma. Figure 2 depicts an example of an archetype for a case of monitoring a blunt traumatic injury.
- In the *clinical data repository,* all information coming from different sources (prehospital, hospital, diagnosis, treatment, etc.) is consolidated.
- The *standardized vocabulary* contains the codes of classification and coding of diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury and/or illness.

Results

MOSPA is an Android ® application that is part of the mobile support environment of the proposed model. This application allows us to record the most important patient information from the MLST system (mechanism of trauma, suspected injury, vital signs, and treatment performed such as CPR, endotracheal tube placement, etc.). The pregnancy status and gestational age, among other prehospital information is also incorporated. In addition, the application cancalculate the type of alert (severe, mild or none) to be sent to hospital personnel and estimate travel time from the scene to the hospital (using Google® maps). The total number of variables that are handled by the application is 29 (according to MLST).

Figure 3 shows an example of the process that allows us to determine the kind of alert that must be sent from the ambulance to the receiving hospital. As shown, the system must analyze the values of each variable of MLST and how they are combined (two or more critical values, two or more

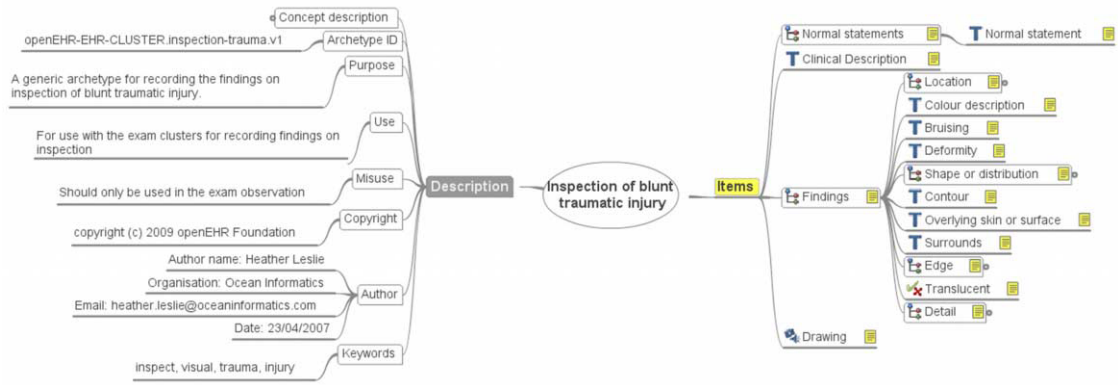


Figure 2 – Partial view of an OpenEHR archetype to perform the inspection of blunt traumatic injury (source: openehr.org).

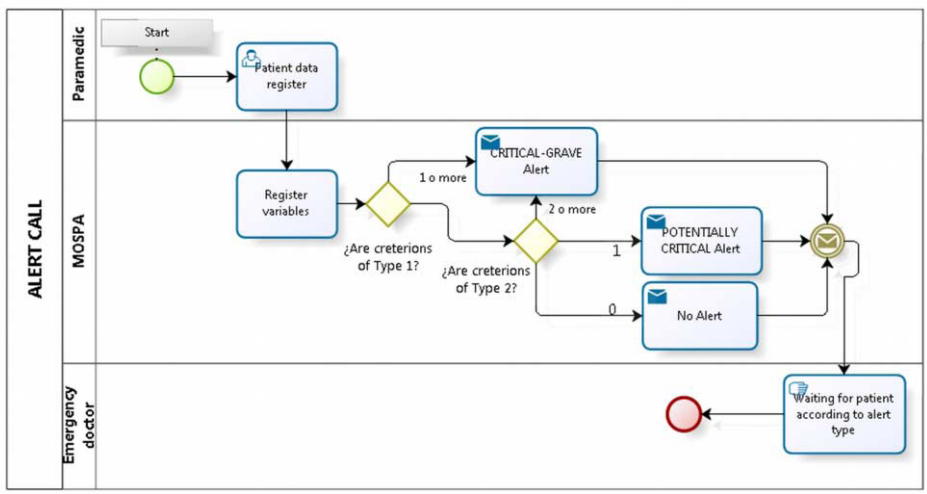


Figure 3– Logical process to determine the kind of alert that will be sent to hospital.

non-critical values, etc.). For example, whenever a patient presents arterial hemorrhage or major bleeding, the system must automatically send a critical alert to the hospital. Likewise, whether a patient has lost more than 500cc of blood and has been unconscious for more than five minutes, the system will send a critical alert.

To verify the tool, we conducted a laboratory pilot test with 32 participants who were trained in the use of the application. As a first step, random searches of trauma cases were generated, and then participants were requested to record such cases in the mobile application. As the participants filled-in all information of the cases, the time taken to record each was measured. This is illustrated in Table 1.

Table 1– Preliminary evaluation results of MOSPA.

Element	Value
Participants	32
Average filing time	59 seconds
Average number of errors comited	7

As we can see, the average time is about one minute, which is an encouraging result, since the participants involved are not professional paramedics, and this allowed us to establish that the application is an excellent alternative to be used in ambulances. To the best of our knowledge, the MLST system is commonly applied for trauma attention by using air transport. The maximum time to enter the information of trauma in these cases must be no more than one minute. Therefore, we consider that for terrain transportation of trauma patients, the time should be similar.

Figure 4a depicts a screenshot of the window that allows entry of information related to pregnancy. Figure 4b showed the route and estimated time of arrival of the ambulance to the receiving hospital.

Discussion

Today, intelligent systems have become a tool that supports the most diverse areas of science, providing various services such as support in decision-making, information discovery, pattern analysis, and artificial vision, among others. Just one of the sciences where such systems can provide important support is medicine because of the wide range of existing knowledge and the complex problems that must dealt with every day [10]. Existing developments are varied and include the application of various techniques related to the fields of engineering and especially with artificial intelligence.

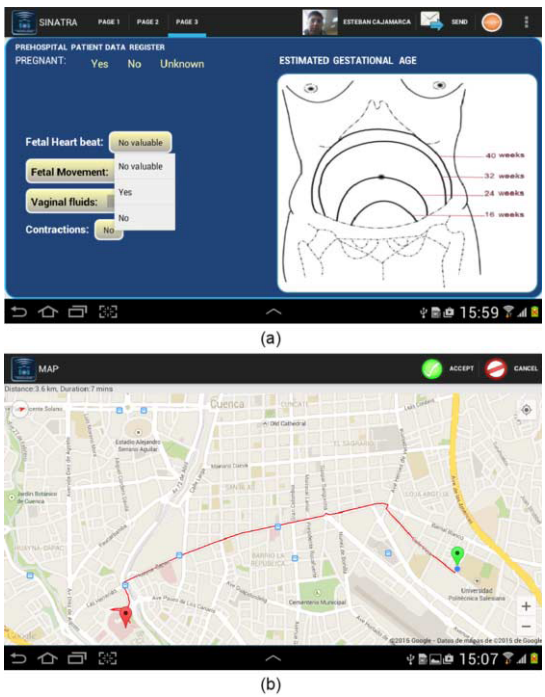


Figure 4 – (a) Screen capture of registration of pregnancy information, (b) Route and estimated time of arrival.

A report presented by Fitzgerar [11] aims to demonstrate that computer systems for decision support can reduce errors during the first 30 minutes of resuscitation when trauma occurs. Also, Han [12] presents a study on the use of expert systems to monitor biometric signals of patients who are at home.

Other areas of study on expert systems to support decision making in medicine focus on the following lines:

- Analysis and prediction of adult patients with intracranial hemorrhagic trauma (HIT): Nishijima conducted a study where an expert system based on rules to determine whether patients over 18 Years with HIT should move to the ICU [12]. The developed system used a set of rules derived from classification and regression trees and obtained promising results.
- Early diagnosis of cervical vascular injuries: Purvis [13] proposed using decision trees for the early detection of blunt cervical vascular injuries. Based on this research, the prevention of such injuries is expected, therefore should be improved outcomes for patients.

As shown, the proposed ecosystem could effectively replace the existing system given that the paramedics (in Cuenca) currently use only radio communication, and the staff hospital can not monitor in a real-time the patient condition. Likewise, currently the ambulance conductor provides an estimation of arrival time, but does not have automatized support to determine the nearest hospital or the best route to follow.

Conclusion

To the best of our knowledge, currently, Ecuador does not have an integrated environment to provide support at the different stages in trauma care. It is crucial to consider all actors involved in this process (doctors, paramedics, support staff, patients and their relatives), and try to optimize the response time and resources required to provide effective trauma care. We believe that the ecosystem described can be utilized as a tool in the process of managing and sharing information related to patient care.

Acknowledgements

This work has been sponsored by the Consorcio Ecuatoriano para el Desarrollo del Internet Avanzado (CEDIA), the Universidad Politécnica Salesiana and the Universidad del Azuay.

References

[1] Organización Mundial de la Salud. Estadísticas Sanitarias Mundials. Francia: Organización Mundial de la Salud, 2011.

[2] Organización Mundial de la Salud. Informe mundial sobre prevención de los traumatismos causados por el tránsito. Ginebral: Organización Mundial de la Salud, 2004.

[3] Astudillo R, Salamea JC, Crespo P, and Salamea P. Trauma: Diez años de experiencia – Hospital Vicente Corral Moscoso. In: Revista Trauma de la Sociedad Ecuatoriana de Trauma, 2007.

[4] Sarmiento G, Sarmiento V, Martínez F, Salamea J. Comportamiento del Proceso Prehospitalario Ante la Llamada de Auxilio por Trauma, CSC-911 2011. In: Panamerican Journal of Trauma, Critical Care & Emergency Surgery, 2012, vol. 1, no. 3, pp. 168–74.

[5] Salamea JC, Sacoto H, Rodas EB. Trauma y emergencias en el sur del Ecuador: Historia, realidad y perspectivas. In: Panamerican Journal of Trauma, Critical Care & Emergency Surgery, 2014, vol. 3, no. 2, pp. 73-75.

[6] González J. Soporte Vital Básico y Avanzado en el Trauma Prehospitalario. 7th Edition. Barcelona: Elsevier - España, 2012.

[7] American College of Surgeons. ATLS Manual del curso para estudiantes. Soporte Vital Avanzado en trauma para médico. 8th ed, Chicago: Librería Médica Emergencias, 2008.

[8] Instituto Nacional de Estadisticas y Censos (INEC). Registro de Defunciones, 2011.

[9] Sistema Integrado de Seguridad ECU 911. Informe de atención mensual centro ECU 911 Austro noviembre 2013, 2013.

[10]Sox HC, Higgins MC, and Owens DK. Medical Decision Making. 2nd Edition. Wiley-Blackwell, 2013.

[11]Fitzgerald M. et al. Trauma resuscitation errors and computer-assisted decision support. In: Archives of Surgery, 2011, vol. 146, no. 2, pp. 218-225.

[12]Han SC, Mirowski L, Jeon SH, Lee GS, Kang BH, and Turner P. Expert Systems and Home-based Telehealth: Exploring a role for MCRDR in enhancing diagnostics. In: International Conference, UCMA, SIA, CCSC, ACIT – 2013, 2013, vol. 22, pp. 121-127.

[13]Purvis D, Aldaghlas T, Trickey AW, Rizzo A, and Sikdar S. A Novel Decision Tree Approach Based on Transcranial Doppler Sonography to Screen for Blunt Cervical Vascular Injuries. In: Journal of Ultrasound in Medicine, 2013, vol. 32, no. 6, pp. 1023-1031.

Address for correspondence

Cristian Timbi-Sisalima

Universidad Politécnica Salesiana, Cuenca – Ecuador

ctimbi@ups.edu.ec

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-333

Quantifying the Activities of Self-quantifiers: Management of Data, Time and Health

Manal Almalki[a], Fernando Martin Sanchez[a], Kathleen Gray[a]

[a] *Health and Biomedical Informatics Centre, University of Melbourne, Melbourne, 3010, Australia*

Abstract

Current self-quantification systems (SQS) are limited in their ability to support the acquisition of health-related information essential for individuals to make informed decisions based on their health status. They do not offer services such as data handling and data aggregation in a single place, and using multiple types of tools for this purpose complicates data and health self-management for self-quantifiers. An online survey was used to elicit information from self-quantifiers about the methods they used to undertake key activities related to health self-management. This paper provides empirical evidence about self-quantifiers' time spent using different data collection, data handling, data analysis, and data sharing tools and draws implications for health self-management activities.

Keywords:

Self-quantification; health self-management; participatory health; consumer health informatics; patient activation.

Introduction

A self-quantification system (SQS) is a tool for capturing personal data and an application for processing that data (e.g., analysing and visualising) according to an individual's objectives. SQS have the potential to track and measure various health aspects, such as sleep, weight, diet, and physical activity, and may offer opportunities for self-quantifiers to change behaviours and attitudes toward health self-care. In one study [1], 69% of US adults said that they tracked themselves, and 46% of those reported changing their behaviour based on the collected data.

There are two types of SQS: A primary SQS collects data directly from the users; health-related examples are MoodPanda for tracking mood and iBGStar for monitoring blood glucose. A secondary SQS can aggregate data delivered by multiple primary tools or apps, to facilitate data analysis; TicTrac, BodyTrack, and Argus are health-related examples [2].

Self-quantification for health consists of two main stages of activities in which a self-quantifier not only collects personal health measurements, but also manages and transform these data into actionable knowledge [2]. The first stage, data management, involves a chain of activites that are necessary to establish, use, and maintain a mapping between the individual's objectives and the information available. The second stage is health management; here the knowledge obtained from self-quantification is activley transformed into the desired health outcomes, a process that may be called health activation [3]. In these activities, the person attempts to establish self-awareness of one's health and functional status to actively engage with maintaining or improving it. Aggregating data from different devices and apps is a major facilitator in gaining a holistic view of one's health-related indicators [4].

To undertake this chain of activities for health data self-management, a self-quantifier may use multiple unconnected primary and secondary SQSs, and also ancillary tools – such as iPhone note apps, Microsoft Excel spreadsheet, or Dropbox. With their data and services scattered among different tools self-quantifiers may devote significant time for these activities. The aim of this paper is to characterise the experience of using combinations of such tools, particularly its impact on self-quantifiers' time and data, and to draw implications for health self-management and health activation.

Methods

Sample and demographic characteristics

An international online survey was conducted from December 2013 to March 2014 to elicit information from adult self-quantifiers, about the methods they used to perform key activities related to health data self-management - data collection, handling, analysis and aggregation, and sharing. Participants were recruited through Quantified Self meetup forums, Twitter, Facebook, and LinkedIn. People who used one or more self-quantification tools or apps as part of their health self-care were included. 103 individuals provided sufficient information for analysis.

Respondents from USA were the highest proportion (60%), followed by Australia (11%). Nearly two thirds (62%) were aged between 20 and 39 years. 75% were male. Almost one third (32.3%) had completed high school or equivalent; the rest had at least a university undergraduate degree. 68% reported good to very good health status. The top three motivations to use SQS were: to know if a certain health-related variable could affect another variable (64%); to find answers to specific questions related to health (62%); to proactively minimise possible future health problems (61%). The Eurostat's ICT (information and communication technology) model [5] was used to classify respondents into high (53%), medium (39%), and low (8%) level of ICT skills.

From the surveyed self-quantifiers, 89% used at least one tangible tracking device for self-quantification whereas 11% used only apps. The most popular primary tangible self-quantification devices were for tracking and quantifying physical activities, sleep, weight, calories burned, and diet: the Fitbit collection (e.g., Fitbit Ultra, Fitbit One, Fitbit Zip, etc.) with nearly 25% of the respondents; followed by Withings

tools (e.g., Withings scale, and Withings activity tracker) with 19%; and Jawbone Up with about 12%. The most popular primary tracking apps were 80Bites (for tracking food consumption), 42Goals (for tracking personal life goals such as quit smoking, lose weight, and reduce coffee consumption), and Moves (records walking, cycling, and running activities through step counter), among approximately 13%, 11%, and 9% of respondents respectively.

The most popular secondary self-quantification tool was BodyTrack (20% of respondents). In addition, two types of ancillary tools were used for collecting data and for handling the generated datasets. The most popular ancillary tools for data collection were diary apps such as Evernote, Microsoft One Note, and iPhone note apps (30% of respondents). Most popular tools for data handling were cloud storage services such as Dropbox, SkyDrive, and Google Drive, collectively used by nearly one quarter (24%) of respondents.

Statistical analysis

Data analysis to address the specific aim of this paper focused on responses to survey questions about the amount of time spent to undertake the key activities of health data self-management (i.e., data collection, handling, analysis and aggregation, and sharing). The survey devoted separate sections to each of the key activities where the participants were asked questions about the type of tools they used (i.e., primary, secondary, or ancillary); their experience in using these tools (i.e., less than 6 months, or 6 months or more); the frequency of tool usage (i.e., daily, weekly, or monthly) for these activities; the number of collected data types (such as blood pressure, weight, mood, coffee consumption, etc.); the methods used to collect these data (i.e., manual or automatic); and their ICT skills level. These questions represent a set of explanatory variables that were fitted in testing the statistical models.

General linear model for continuous outcomes, and logistic regression for binary outcomes, were conducted by using SPSS 22 Mac version. To construct these models, a manual backwards selection approach was used. We started with set of candidate explanatory variables and eliminated variables with large p-values one at a time. We stopped once all the remaining variables had relatively small p-values. It is worth noting that building these statistical models did not use only a simple analysis of concepts; rather, it involved exploring a more complex concept through testing for a two-way interaction. This two-way interaction is often thought of as a relationship between an explanatory variable and outcome variable that is moderated by a third explanatory variable. Once a two-way interaction between two varables was found, the main effects of the explanatory variables are not discussed given the presence of these interactions. Bootstrapping results were examined to check the robustness of our models. Also, SPSS 22 was used to run descriptive statistics, frequencies, and cross-tabulations to determine related demographics.

Due to the sample size limitation , we combined the two categories of medium and low ICT skills into one group (low to medium level). Also, the ICT skills level variable could not be fitted with other explanatory variables in a single logistic regression model. Adding this variable increases the number of categorical variables in the model which can result in empty cells and problems in model fitting. Therefore, two logistic regression models were built for these cases, one based on the ICT skills variable and the other including all other variables.

Results

Combined use of primary, secondary, and ancillary tools

No survey respondents were using only primary SQS for health data self-management. Nearly a third used all types of tools, i.e. primary, secondary, and ancillary. Nearly 80% used three primary tools to collect health related data; however, 51% stated that their primary tools did not offer the kind of data they needed and therefore had to use ancillary tools to collect the desired data.

75% of respondents used another type of ancillary tool for handling the collected data from these primary SQS, for example, to organise the collected data into folders, import or export data, update data or files. 70% of respondents were performing most of the related tasks manually. Thus, over two thirds (68%) had lost files permanently, and over three quarters (77%) had failed to locate old files.

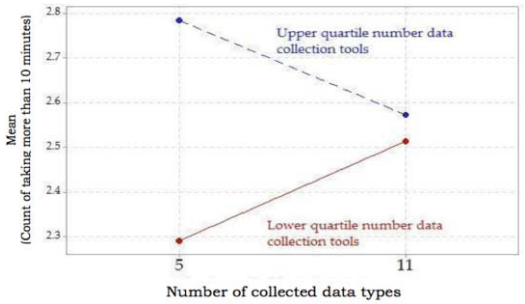
75% of respondents used secondary SQS to aggregate the data generated from the primary SQS. In over a third (34%) of these cases the secondary tools did not connect automatically with the primary tools, requiring users to upload the datasets manually into the secondary SQS.

Time devoted to data collection

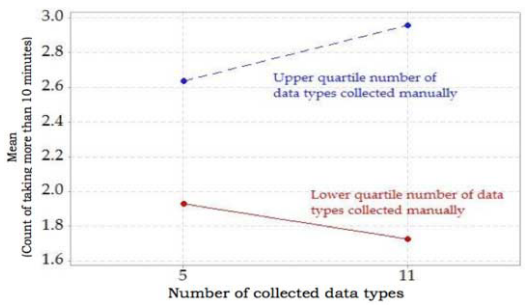
Survey respondents who used multiple primary and ancillary tools to collect the required data spent more time on average than those who used fewer tools. The data collection time was calculated by counting the cases that took over 10 minutes to acquire the needed data using these tools. For representation simplicity, we refer to this count with the symbol CTM (Count of Taking more than 10 Minutes). To measure the impact of using these tools on CTM, we identified a set of factors (explanatory variables) that had a statistically significant effect on the mean of CTM. These factors were: the number of primary and ancillary tools used for data collection; the number of collected data types; the method used to collect these data; and the ICT skills level. Three two-way interactions between these variables existed in this model (see Table 1 model number 1). The first two-way interaction was between the number of tools used and the number of data types collected (by the primary and ancillary tools used). The second interaction was between the number of data types collected and the method used to collect these data(manual or automatic). The last interaction was the number of collected data types and the ICT skills level. For reporting the results of these two-way interactions, the main effects of the explanatory variables that involevd in these interaction are not presented in the Table 1.

In a general linear model with the set of these factors and their interactions (Table 1 model 1), the effect of the number of tools on the mean of CTM was greater for users who collected relatively fewer types of data. For example, if the number of data types was five, and three tools (lower quartile) were used to collect these data, then the average CTM was 2.3; in comparison, adding one more tool (upper quartile) to collect these five data types increased the mean of CTM to 2.8 (Figure 1-A). The mean of CTM increased with the rise in the number of data collected manually. For example, if the number of data types was eleven and all were collected automatically (lower quartile), then the mean of CTM to collect these data was 1.7; however, if two (upper quartile) out of these eleven data types were collected manually, then the mean of CTM activity was 2.9 (see Figure 1-B). Similarly, the

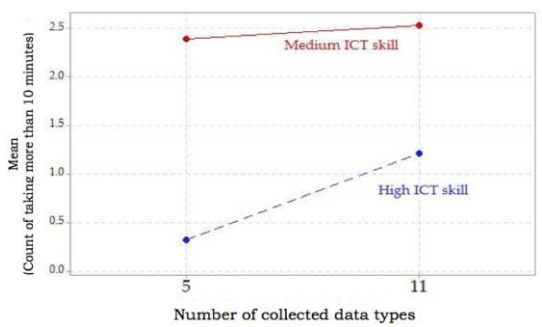
mean difference in time taken for data collection between users with low to medium ICT skills and users with high ICT skills was statistically significant (mean difference=1.65). On average, for collecting five types of data, people with high level ICT skills took less time (0.4) relative to those with low to medium ICT skills (2.4) (Figure 1-C).



(A)



(B)



(C)

Figure 1 - Effects of three different two-way interactions on the mean of count of taking more than 10 minutes

Time devoted to data handling

Using ancillary tools for data handling increased the time needed to complete the activity. To measure the tools' impact, we identified three factors that had a statistically significant effect on the odds ratio of the duration of data handling. These factors were: using ancillary tools for data handling; the frequency the tool usage; and ICT skills level. Two models were built to examine this, due to model fitting problems as explained previously. One model included the first two factors and the other included the ICT skills level.

In the first logistic regression model, the odds of taking more than 10 minutes to complete the data handling activity were statistically significant - six times higher among users of ancillary tools relative to non-users (Table 1 model 2). With growing data volume, among people who used the tool less frequently (on a weekly or monthly basis), the odds of a longer duration of data handling activity were four times higher than among those who used the tool on a daily basis.

In the second logistic regression model (Table 1 model 3), the effect of having low to medium ICT skills on the odds of the data handling duration was significant - 18 times higher compared to users with high ICT skills.

Time devoted to data aggregation and analysis

Self-quantifiers who used secondary SQS took longer times to perform this activity. To measure the impact of using these tools, we identified four factors that had statistically significant effects on the odds ratio of the longer duration of data analysis. These factors were: using secondary SQS; the number of collected data types; the method used to collect these data manually; and ICT skills level. There was a two-way interaction between the number of collected data types and the method used to collect the data. These variables were fitted in two models (Table 1 model 4). One model included the first three factors and the two-way interaction, and the other included the ICT skills level.

In the first logistic regression model, the odds of taking more than 10 minutes to complete this activity were five times higher among secondary SQS users relative to the non-users. People who manually collected multiple types of data took a longer time to analyse these data. The longer duration odds increased with each additional type of data collected manually (see Figure 2). For example, if the number of data types was eleven (third quartile) and all were collected automatically (Manual data collection=0), the odds ratio of data analysis duration (taking more than 10 minutes) was 2.04; in comparison, if two out of these eleven data types were collected manually (Manual data collection=2), then the odds ratio was higher (Odds ratio=6.79) (Figure 2).

In the second logistic regression model (Table 1 model 5), the effect of having low to medium ICT skills on the data analysis duration odds was significant 17.6 times higher compared to users with high ICT skills.

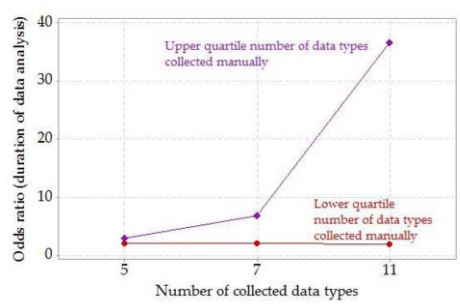


Figure 2 – Effect of the number of manually collected data types on the odds ratio of data analysis duration

Table 1 – Results from statistical models

Factors	Description of Estimate	Estimate Value	*P*	95% CI
Model (1): General linear model for the duration of data collection activity				
First interaction	Difference in regression slope between 'number of data types' and 'number of tools used to collect these data'	-0.073	0.012	-0.129, -0.017
Second interaction	Difference in regression slope between 'number of data types' and 'method used to collect these data (Manual)'	0.044	0.041	0.002, 0.086
Third interaction	Difference in regression slope between group with 'high' ICT skills and group with 'low to medium' ICT skills	0.126	0.024	0.017, 0.236
Model (2): Multiple binary logistic regression model for the duration of data handling activity				
Use ancillary tool	Odds ratio of taking more than 10 minutes in users compared to non-users groups	6.368	0.001	2.268, 17.876
Frequency of use	Odds ratio of taking more than 10 minutes in users who used this tool less frequently (weekly or monthly) compared to frequent users (daily)	4.061	0.004	1.556, 10.600
Model (3): Simple binary logistic regression model for the duration of data handling activity				
ICT skills	Odds ratio of taking more than 10 minutes in users with low to medium ICT skills compared to high level users	18.000	< 0.001	4.987, 64.971
Model (4): Multiple binary logistic regression model for the duration of data analysis activity				
Use secondary SQS	Odds ratio of taking more than 10 minutes in users compared to non-users	5.362	0.001	1.905, 15.091
Number of manually collected data (interaction)	Odds ratio of taking more than 10 minutes with increase in the number of data types collected manually	1.244	0.016	1.042, 1.486
Model (5): Simple binary logistic regression model for the duration of data analysis activity				
ICT skills	Odds ratio of taking more than 10 minutes for users with low to medium ICT skills compared to high level users	17.600	< 0.001	4.872, 63.575
Model (6): Multiple binary logistic regression model for the duration of data sharing activity				
Use secondary SQS	Odds ratio of taking more than 10 minutes in users compared to non-users	31.829	0.003	3.282, 308.7
Number of data collection tools (a)	Odds ratio of taking more than 10 minutes when (a) increased by $\frac{a+1}{a}$	0.283	0.012	0.106, 0.758
Method used for data collection (Manual) (b)	Odds ratio of taking more than 10 minutes when (b) increased by $\frac{b+1}{b}$	3.281	0.003	1.512, 7.122
Experience in using the tool (c)	Odds ratio of taking more than 10 minutes when (c) increased by $\frac{c+1}{c}$	2.667	0.023	1.146, 6.207
Model (7): Simple binary logistic regression model for the duration of data sharing activity				
ICT skills	Odds ratio of taking more than 10 minutes in users with low to medium ICT skills compared to high level users	50.667	< 0.001	11.567, 221.936

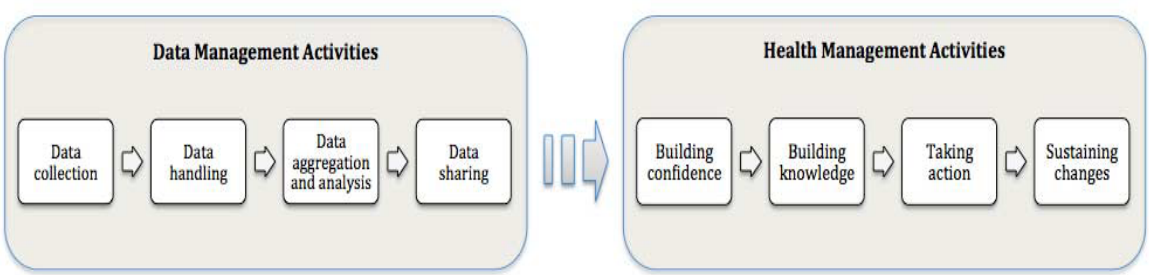


Figure 3 – The gap between data management activities and health management activities

Time devoted to data sharing

Most self-quantifiers in our study (73%) shared their data and findings with other people using primary and secondary SQS. Multiple primary tools usage was found to reduce the time needed to perform this activity; however, using secondary SQS increased the duration of this activity. To measure the impact of using these tools on the odds of the increased duration of data sharing we built two models. In the first logistic regression model, the variables were: using a secondary tool; the number of tools used for data collection; the method used for data collection (e.g., manual or automatica); the degree of experience using these tools; and ICT skills level.

The odds of taking more than 10 minutes to complete the data sharing activity increased 32 times in the secondary tools users compared to the non-users. On the other hand, the odds for longer duration reduced somewhat (0.3 times) for each additional primary tool the self-quantifiers used to share the data (Table 1 model 6). The range of the predicted estimates (confidence interval (CI)) of the odds of longer duration of data sharing when using secondary tools was wider due to limitation in the sample size. The odds of data sharing activity taking longer increased three times with additional types of data collected manually. The odds of longer duration also increased when self-quantifiers' experience in using the tools increased (Odds ratio=2.6).

In the second logistic regression model, the effect of having low to medium ICT skills on the odds of data sharing activity taking longer was 50 times higher compared to users with high ICT skills (Table 1 model 7).

Discussion

This study offers a detailed understanding of the SQS usage experience for health data self-management activities. The results show that managing day-to-day health requires individuals to interact directly with the chosen tools to collect, handle, analyse, or share the generated data. The time needed to go through each of these activities towards achieving one's desired health outcomes depends on the tools' ability to provide the required information and services. Having high ICT skills can slightly reduce the amount of time self-quantifiers have to devote to these activities. These findings are consistent with previous research studies [4, 6, 7].

Currently, few single primary self-quantification systems can support all information needs of a user. This limitation could generate a gap between the two critical stages of health self-quantification (Figure 3). We argue that as the gap gets wider, a self-quantifier has to focus more on managing data instead of managing health; thereby, facing information and time limitations on taking an active role to turn the collected data into knowledge, and subsequently take informed actions. Additionly, using multiple types of tools could diminish the data quality and increase the risk of losing data (files). The senario gets worse when the individuals do not have high ICT skills.

Conclusion

Using multiple types of tools for the purpose of health self-quantification complicates the health data management activities. In particular, it has implications for data and time management that pose challenges for achieving good health outcomes. We suggest that a framework representing and describing all key activities in health self-quantification could be beneficial. Application developers and health researchers could use this framework to improve the design and evaluation of health self-quantification solutions. Healthcare providers could use this framework to determine the applicability of these solutions and plan their implementation, based on the expected outcomes. Thus our current research focuses on the development of a framework, based on the theories of behaviour changes and human interactions with SQS tools, allowing a comprehensive view of all the essential activities and components of health self-quantification practice. We are also conducting further empirical research on the relationship between self-quantifiers' data management activities and health activation, to better understand the factors involved in their ability to achieve desired health objectives.

Acknowledgments

Funding for this project was provided by the University of Melbourne's Institute for a Broadband-Enabled Society (IBES). Almalki's PhD research is supported by a scholarship from Jazan University and a top-up scholarship from IBES.

References

[1] Fox S, Duggan M. Tracking for Health. *Pew Research Center's Internet & American Life Project. 2013*. Available from: http://www.pewinternet.org/2013/01/28/tracking-for-health

[2] Almalki M, Gray K, Martin-Sanchez F. The Use of Self-Quantification Systems for Personal Health Information: Big Data Management Activities and Prospects. *Health Information Science and Systems HISS*; Australia, Melbourne. 2014.

[3] Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers. *Health Services Research*. 2004;39(4p1):1005-26.

[4] Li I. Personal Informatics and Context: Using Context to Reveal Factors that Affect Behavior. *Journal of Ambient Intelligence and Smart Environments*. 2012;4(1):71-2.

[5] Eurostat. 2011. Available from: https://circabc.europa.eu/sd/a/2ea7b600-242c-46db-9e4f-afb25e03ba87/Questionnaire%20HH%202011%20v.3.1.pdf

[6] Jones W. Personal Information Management. *Annual Review of Information Science and Technology*. 2007;41(1):453-504.

[7] Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Canada, Toronto. 2014;1143-1152.

Address for correspondence

Manal Almalki. University of Melbourne, 3010, Australia. E-mail: malmalki@student.unimelb.edu.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-338

Mobile Usability Testing in Healthcare: Methodological Approaches

Elizabeth M. Borycki[a], Helen Monkman[a], Janessa Griffith[a], Andre W. Kushniruk[a]

[a] *School of Health Information Science, University of Victoria, Victoria, BC, Canada*

Abstract

The use of mobile devices and healthcare applications is increasing exponentially worldwide. This has lead to the need for the healthcare industry to develop a better understanding of the impact of the usability of mobile software and hardware upon consumer and health professional adoption and use of these technologies. There are many methodological approaches that can be employed in conducting usability evaluation of mobile technologies. More obtrusive approaches to collecting study data may lead to changes in study participant behaviour, leading to study results that are less consistent with how the technologies will be used in the real-world. Alternatively, less obstrusive methods used in evaluating the usability of mobile software and hardware in-situ and laboratory settings can lead to less detailed information being collected about how an individual interacts with both the software and hardware. In this paper we review and discuss several innovative mobile usability evaluation methods on a contiuum from least to most obtrusive and their effects on the quality of the usability data collected. The strengths and limitations of methods are also discussed.

Keywords:
Usability; Mobile phone; Mobile health; In-situ; Healthcare; Consumer informatics; Clinical informatics; Human factors.

Introduction

The use of mobile devices and healthcare applications among consumers and health professionals is rapidly increasing [1, 2]. Athough there has been an exponential rise in the use of mobile devices in conjunction with m-health applications, few researchers have explored the methodological approaches and issues encountered when conducting in-situ usability testing (i.e. usability testing conducted in the setting of use) in environments where healthcare activities take place [1-3]. In this paper the authors will outline several innovative approaches to conducting mobile usability testing of mobile healthcare applications and their devices. This work will also include a discussion of the strengths and limitations of each of these approaches in the context of quality of data collection.

Review of the Literature

With the advent of mobile health or m-health, more and more health professionals and consumers are using mobile devices along with healthcare software applications (e.g. mobile devices that provide diet and excerise advice). Consumers are using these devices to self-manage their wellness activities (e.g. diet and exercise) as well as chronic illnesses (e.g. chronic obstructive pulmonary disease, hypertension). Health professionals are also using these devices to review, communicate and undertake health care activities (e.g. prescribing medications, reviewing clinical guidelines) [4]. Yet, even as the use of these devices has grown in the consumer and health professional space, there has been less attention given to evaluating the usability of mobile devices and their software applications in healthcare [5]. General issues in the usability testing of mobile devices and their software applications have been described by a number of authors, including Nielsen and Budiu [6], Weiss [7] and others such as Pearrow [8]. However, these works have focused on conducting laboratory-based studies with fixed recording of users using mobile applications in laboratory settings [6-8]. In addition, research is needed for exploring the usabilty of complex eHealth applications that involve understanding their use and usability under both laboratory and real-world conditions in healthcare contexts (e.g. hospitals, clinics) or in contexts where health is being managed (e.g. at home, at the gym or at work) [3-5]. We describe and compare a number of methods that we have employed in in-situ usability testing of mobile devices and healthcare applications in this paper.

Review of Existing Materials and Methods

We have employed several mobile usability testing approaches. The approaches also vary to the extent they are invisible or unobstrusive to the end user of the mobile device and the healthcare applications being tested. It has been our experience that some users may find being audio and video recorded to be obstrustive. Obtrusiveness may lead some users to modify their performance of tasks, improve their behavour in order to please the usability researcher, and/or limit their negative comments about the software features, functions and layout in order to “please” or to “not hurt the feelings” of the researcher. This is known as the Hawthorne effect. The Hawthorne effect can compromise the quality or representativeness of the study results. In our work, we have identifed that less obstrusive methods of audio and video recording user interactions with hardware and software devices leads to the collection of data that is more representative of real-world settings in terms of participant behaviour [9]. Yet, in the process of reducing the obstrusiveness of recording techniques, the quality of the data collected is diminished in some cases (such as losing data about finger and hand gestures) and may limit the researchers ability to identify all potential usability issues [10]. We have found that obtrusivness of recording methods exists on a continuum. There are a number of tradeoffs that exist when selecting an approach to usability testing where obtrusiveness is concerned. On one end of the continuum of obtrusivness of usability appraoches, the mobile device the user interacts with

is also used to audio record user's "think aloud" verbalizations and video record healthcare software application screens (directly to the mobile device) as the user is interacting with the software. On the other end of the continuum, the usability researcher uses external (i.e. visible to the user and researcher) audio and video recording devices to collect user interactions (including finger and hand gestures) with the mobile device and the eHealth application under study. In between these two points on the continuum, we have differing levels of obtrusiveness of audio and video recording software and devices where user interaction with the eHealth application and mobile device is concerned. As the obstrusiveness of the recording methods increase so does their influence on participant behaviour which may become less representative with the user describing fewer usability problems in an effort to please the researcher.

Figure 1 –Continuum of Obtrusiveness

In addition, the approaches vary in the extent to which they can be effectively used to collect data in laboratory-based and in-situ settings. All of the approaches can be used in a laboratory settting (i.e. a room set aside for usability testing). Many of the approaches can also be used in-situ, depending on the tasks that the users will be asked to perform and the level of user mobility that is required to perform the task; for example, a physician user may be asked to perform a task typically done while sitting at a desk such as ePrescribing versus the nurse user having to move, bend over and stand while performing a task such as taking and recording a patient's vital signs.

Usability Testing with the Device as Screen and Audio Recorder

Usability testing with a mobile device, where the device also audio records user verbalizations and records the screens the user is interacting with, is the least obstrusive of these approaches (see Figure 1, left side of the continuum). The mobile device provides the user with access to mobile eHealth applications and is also used to collect verbal and screen recording data (see the left side of Figure 2). In this way, the usability researcher can use the device to record the participant's "think aloud" verbalizations and video record the screens the user moves through while performing the task (see the right side of Figure 2 for the Data View i.e. the view the researcher sees when playing back the recording for analysis).

Figure 2 –Device as Screen and Audio Recorder

There are a number of advantages to using this approach. In terms of hardware, only the mobile device itself is used, so the cost of conducting such usability tests is low: limited to the cost of the mobile device and the audio and screen recording application. The usability researcher installs the eHealth application, and the audio and screen recording application (e.g. Screen Recorder by ToySoft® on the mobile phone). The audio and screen recording application is used to record the user's verbalizations and to record the eHealth software application screens as the user moves through the application (see Figure 2). The approach affords the usability researcher and user high levels of portability. The use of a single mobile device with a healthcare software application and audio/screen recorder allows the user to move freely in a laboratory or an in-situ setting, performing tasks that require walking or other types of movement (e.g. taking a picture of a surgical wound to document healing in the patient record). From an in-situ perspective, the usability researcher can easily bring the device into the users' environment, begin the recording, and the user can use the device as he or she would in their healthcare setting. The ecological validity of this approach is very high as the environment, software and hardware are representative of what is used in the real-world. In addition to this, for studies of use of mobile eHealth applications in collaborative settings, each user may have the recording software deployed on their own mobile device, allowing for multiple recordings (that can later be synchronized during the analysis).

The one significant disadvantage associated with using this approach is that the users' finger, stylus or hand gestures, are not recorded so it is unclear as to exactly what the user is touching on the eHealth application interface (e.g. a button several times before the software responds), as this is not recorded (see right side of Figure 2). In addition to this, the impact of the mobile device on user activities is not fully known. In our studies, features of the hardware device may have an impact upon users' perceived usability of the software and hardware and this information is not fully recorded as there is no external view of the participant handling the mobile device. It must also be noted that not all mobile devices have recording software that can be effectively used to capture the screens that the user is moving through and many mobile devices will not have sufficient room to store large files used in screen recording. Ideally, the device the consumer plans to use or the health care organization has mandated should be selected for the testing to ensure ecological validity and representativeness of the data [10,11]. Pilot testing [10] of the study data collection methods is recommended, including testing the mobile health application, mobile device and recording software to ensure no interactions between differing software and hardware lead to audio and video data being lost.

Usability Testing by Mirroring to a Computer

Along these lines, in cases where a mobile device has insufficient memory to store audio and screen recordings from the usability test or in cases where audio and screen recording applications are of insufficient quality to provide high quality recordings, devices can be mirrored to a computer (e.g. transmitting iPhone® or Samsung Phone® screens to a computer). Usability testing by mirroring a device screen to a computer or laptop computer remotely and then using audio and screen recording software installed on the computer can be done effectively to record audio and video data. In other studies, we have found that computer screens (including a mirror image of the mobile health or eHealth application

display) can be recorded using software installed on a computer. The approach is of higher cost as a laptop must be purchased; however, costs can be limited by using free software that can be downloaded off the web such as HyperCam® to record mobile device screens.

We recommend pilot testing this approach prior to collecting usability data [10]. Care must be taken by the usability researcher to ensure that the mirror image of the device is of sufficient quality to provide video data, and a portable external microphone is carried by the user that connects the user and their device to the laptop with its audio and video recording software so as to fully capture think aloud verbalizations in conjunction with video data of the user moving through the eHealth software application screens (see Figure 3).

Figure 3 –Mirroring to a Computer

The approach is more obtrusive than *usability testing with the device as screen and audio recorder* as the user is aware of the usability researcher reviewing the recording using a laptop and the user is physically wearing a portable microphone. However, the user is able to move freely when using the device (provided it is sufficiently close enough to the laptop for the computer screens and audio verbalizations to be recorded). *Mirroring to a laptop* does have a number of drawbacks – cost increases to conduct the test as the user now needs to use a laptop. In addition to this, less information is gathered as the researcher is unable to record user finger, stylus and hand gestures when touching eHealth application screens as they interact with the healthcare software. However, this approach takes care of the problem of not having enough space to record files on the mobile device, as the recordings are stored on the computer the mobile screens are mirrored to.

Usability Testing Using a Headcam

Head Cams can also be used to conduct mobile usability testing. Here, the individual wears a camera that is fixed to an adjustable strap that is worn on the forehead. As the user interacts with the mobile device, video and audio data are recorded by the headcam including finger and hand gestures. Usability researchers must take care when selecting a headcam when conducting this type of research. Headcams need to have audio equipment built in. Such headcams are currently available for purchase (e.g. the goPro® camera) with a high resolution and frame rate, thereby allowing for the usability researcher to capture user interactions with the mobile device and healthcare software application. The device is more obtrusive with it being worn by the user on their head. Alternatively, the device does afford the user participant greater mobility and it is possible to observe what the user is focusing on in terms of the mobile healthcare software. The user must be trained to position their head to ensure the headcam captures the mobile device screens and their finger/hand gestures for recording (see Figure 4).

Figure 4 –Using a Headcam

Here, the headcam improves the quality of the data collection by allowing the usability researcher to record how the user is interacting with the device (i.e. we can see the user touching differing user interface features and functions using a stylus). The approach is also reliant on only one device (the head mounted camera) so it requires limited equipment.

Usability Testing Using Glasses

We are currently adapting our approaches to using glasses that record audio and video for usability testing of healthcare applications. Our initial investigation suggests that such eyewear can be effectively used to collect information about what participants are viewing in addition to hand, finger or stylus movements (See Figure 5).

Figure 5 –Using Glasses

The advantage of the approach is that the researcher can view the world as the participant views it, and the participant is able to physically move while wearing the glasses. Our initial research has found that careful consideration must be taken when selecting the type of glasses that will be used in usability testing; for example, glasses should be selected with the ability to record audio and video data (as not all glasses record audio data). Attention must also be given to ensuring the glasses have sufficient storage space to collect the data that is being sought and there is a need to determine at what points downloading should take place. The glasses themselves must also be considered in terms of their usability and ergonomics. Some eyewears are difficult to position on the nose to fully record activities. Other glasses have been reported by users as difficult to wear. Some users find wearing glasses that can video and audio record to be irritating to the nose. As well, some glasses that record audio and video data do not respond well to participants head movements (i.e. the video may become choppy and may not

fully capture what the participant is looking at). Some that are able to record audio and video data may also have a headup display. The headup display provides information to the participant in the form of a transparent display. The user sees data in their visual field and does not have to look away in order to review the information, much as a pilot is able to continue looking forward at the data typically found on an instrument panel rather than needing to look down to view his or her instrument panel while flying an airplane [12]. The usability researcher must take into consideration if the headup display supports or distracts the user from focusing on the mobile device and healthcare application. Distractions may diminish the quality of the participants "think aloud" as they focus on the headup display instead of the mobile device. Lastly, some individuals use eye glasses for vision correction. Glasses used for audio and video recording usability data cannot correct a participant's vision. Therefore, some participants may be excluded from being part of this study. In summary, careful consideration must be taken by the usability researcher when procuring and pilot testing [10] glasses that record video and audio data for the purpose of collecting usability data about mobile device and software application use.

Usability Testing Using a Document Camera

Mobile usability testing can also be done using newer, more portable document cameras [14]. Products such as the Hue® document camera can be easily taken to a in-situ setting and used for usability testing. The quality of document cameras has improved significantly over the past few years. They are smaller in size (approximately 25 cm long and 10 cm wide), have built in microphones, 10x the zoom function, can be adjusted to focus on any device, and can be plugged into a computer where the images and audio it captures is easily recorded (see Figure 6).

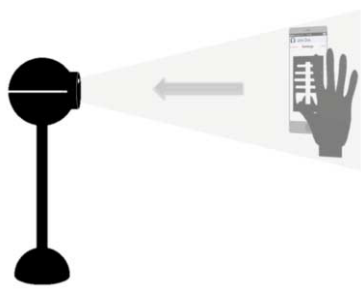


Figure 6 –Using a Document Camera

Document cameras offer a number of advantages. The smaller size of the camera allows for ease of portability and set-up in real-world settings, and the built in microphone no longer requires that the usability tester carry a microphone for capturing think aloud verbalizations [14]. A document camera has a number of advantages over the *use of the device alone for usability tests*, *mirroring to a computer* or *mirroring to a video projector*. Document cameras allow for recording of device screens, user hand and finger interactions with the software application as well as the mobile device. The document camera can be positioned to capture user interactions with the hardware and software. Some document cameras can be clipped to the device using plastic clips [14]. In capturing user interactions, the researcher is better able to link user activities involving the device to screen and audio recordings; for example, you can see what the user is touching with a stylus or their finger on the software application and this information is recorded. This approach does, however, require the mobile device to be located in a relatively fixed position, so its screen can be recorded by the camera (which makes it useful for laboratory style studies but less useful for studies where the user is moving around in a real or realistic environment). The methodology, although more intrusive, does provide additional insights as to what the user is doing with the software and the device.

Usability Testing Mirroring to a Projector

Early studies of mobile applications in healthcare focused on varying aspects of mobile device and software usability. For example, Kushniruk and colleagues found that the small screen size of a Personal Digital Assistant (PDA) could lead some users to make an error [11]. In this study, a statistical relationship between usability problems and medication errors was found by analyzing physician interactions with a mobile ePrescribing application installed on a PDA. The approach involved having end users of a prescription writing application (for use on a palm pilot) projected to a screen using an attachment called Presenter-to-go (which allows for connection of the mobile device to a data projector). Then, the usability researcher used a video camera to record both the computer screens and the audio of users while they interact with the ePrescribing application (see Figure 7).

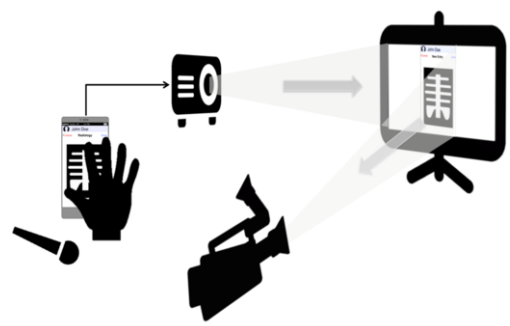


Figure 7 –Mirroring to a Projector

This approach allows for recording of mobile device screens and audio of the participants' "think aloud" verbalization, but requires a cable from the mobile device to the data projector. Usability testing mirroring to a projector is more obstrusive on the continuum of approaches, placing this method more to the right side of the continuum. Here, the user may become more aware of the presence of the video projector and the video camera. Also, the user is tethered to the projector and this limits what the user can do in terms of interacting with their environment (as they are limited by a cord), and less information is available about how the user is interacting with the mobile device (i.e. touching the screen). Using this approach requires more equipment, but as most of the equipment is normally available in most organizations it may be cost effective. Futhermore, the approach does not require installing software to record directly to the mobile device [11].

Discussion and Conclusion

When conducting usability tests involving devices, usability researchers must consider a number of differing issues, obstrusiveness is one of them. The level of obtrusiveness is

important to consider as participants may may modify their behaviours when performing study tasks as they are aware of the presence of equipment. In our work we have found that less obstrusive approaches to gathering usability data improve the ecological validity of the study, while at the same time reducing the likelihood of there being a Hawthorne effect.

Some less obtrusive approaches allow the user to move the mobile device and perform work activities more freely, but may not allow one to record stylus, finger or hand motions on the mobile device screen nor do they collect data about how the mobile device is held or used by the study participant (i.e. *usability testing with device as screen and audio recorder, mirroring to a laptop, mirroring to a projector*). Some more obstrusive recording approaches such as *using a document camera* may limit what the user can physicially do with a device, but will allow for recording of some more sedentary tasks such as entering medications into an ePrescribing system [11]. The headcam or glasses, although still obtrusive, may offer an advantage in that they allow for recording of user audio and video data involving hand gestures and finger touching where the healthcare software application and mobile device are concerned. Along these lines, some participants may find headcams and glasses to be a burden even though the recording method is less obtrusive than others used to collect data mentioned in this paper. There are advantages and disadvantages that need to be considered when planning for a study. Attention also needs to be paid to the types of tasks users will be asked to perform when identifying the best approach to data collection. For example, if the intent of the research is to study physicians using a Smartphone application that is used to for ePrescribing and there is interest in learning about how the device and the eHealth application work together, then using a document camera may be a solution.

Other factors may also influence the researchers' use of recording approaches. Availablity of screen and audio recording software for a Smartphone will determine if *usability testing with the mobile phone as recording device* can be employed as a methodology. Issues such as sufficient storage space for screen and audio recordings to be collected using a Smartphone is a concern. Some Smartphones do not have sufficient storage space to collect video and audio data from a usability session. As well, even if the Smartphone has sufficient storage space for usability recordings, the researcher needs to be able to determine at what point to download data to a computer or external hard drive from the device to free up space for continued recording. Use of laptops and video cameras to record data using mirroring overcomes some of these storage limitations. Cost is another aspect of mobile usability testing that needs to be considered. Using a Smartphone along with low cost screen and audio recording software is cost effective. The use of a headcam, glasses, laptop or a projector adds to the cost of conducting mobile healthcare research (if the equipment is not already in an organization).

In making these choices, researchers need to consider the strengths and limitations of each of these recording methods and their impacts on user behaviour to fully capture the data that answer the researcher's questions. From our work, we have also learned that pilot testing devices, study procedures and data collection approaches (audio and video) is key to avoiding any possible loss of data [10,13]. Also, a fullsome understanding of the tasks and activities that the user will be asked to perform will also influence selection of recording devices. For example, if the participant is asked to enter information into their mobile device while performing a physical activity (e.g. giving medication), the equipment that will be selected for use will differ from tasks where the user may be sitting (e.g. ePrescribing).

In summary, obtrusiveness influences ecological validity and the quality of the study results. Less obstrusive methods of capturing usability data may lead to poorer quality data. Consideration of study design is key when selecting from a range of approaches and taking into account whether the recording devices are obtrusive. This is particularly important in healthcare where there are many contexts and settings of use for mobile applications and devices (e.g. tablet and SmartPhone). In our studies, we have typically employed more than one of the methods described in this paper in combination. Such a multi-method approach may be useful when studies are used to assess both detailed and fine-grained user interactions as well as understanding how mobile applications fit into complex healthcare workflows.

References

[1] Internet World Stats 2010.

[2] Statistics Canada. Internet use by individuals. 2010:http://www40.statcan.gc.ca/101/cst01/comm29a-eng.htm accessed on 11-10-2010.

[3] Kushniruk AW, Borycki EM, Kuwata S, Kannry J. Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments. Studies in health technology and informatics 2010: 169: 915-9.

[4] Fox S, Duggan M. Mobile health 2012. Pew Research Center's Internet x0026 American Life Project, 2012.

[5] Varshney U. Mobile health: Four emerging themes of research. Decision Support Systems 2014: 66: 20-35.

[6] Nielsen J, Budiu R. Mobile usability. New Riders: Indianapolis, 2013.

[7] Weiss S. Handheld usability. John Wiley and Sons: West Sussex, England, 2002.

[8] Pearrow M. The wireless web usability handbook. Charles River Media: Hingham Massachusetts, 2002.

[9] Obrenovic Z. The Hawthrone studies and their relevance to HCI research. Interactions 2014: 221(6): 46-51.

[10] Borycki EM, Kushniruk AW, Kuwata S, Kannry J. Use of Simulation Approaches in the study of clinician Workflow. In AMIA Annual Symposium Proceedings, AMIA 2006: 61-5.

[11] Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application. International journal of medical informatics 2005: 74(7): 519-26.

[12] Weintraub DJ. Human factors issues in head-up display design: The book of HUD (No. SOAR-92-2). Dayton University: Ohio University of Research,1992.

[13] Melbourne B. Recording mobile device usability tests guerilla style. 2014 http://www.thoughtworks.com/insights/blog/recording-mobile-device-usability-testing-sessions-%E2%80%93-guerrilla-style.

Address for correspondence

Elizabeth Borycki: emb@uvic.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-343

Making an APPropriate Care Program for Indigenous Cardiac Disease: Customization of an Existing Cardiac Rehabilitation Program

[a]DanaKai Bradford, [a]David Hansen, [a]Mohan Karunanithi

[a]Australian eHealth Research Centre, Commonwealth Scientific and Industrial Organisation (CSIRO), Australia

Abstract

Cardiovascular disease is a major health problem for all Australians and is the leading cause of death in Aboriginal and Torres Strait Islanders. In 2010, more than 50% of all heart attack deaths were due to repeated events. Cardiac rehabilitation programs have been proven to be effective in preventing the recurrence of cardiac events and readmission to hospitals. There are however, many barriers to the use of these programs. To address these barriers, CSIRO developed an IT enabled cardiac rehabilitation program delivered by mobile phone through a smartphone app and succesfully trialed it in an urban general population. If these results can be replicated in Indigenous populations, the program has the potential to significantly improve life expectancy and help close the gap in health outcomes. The challenge described in this paper is customizing the existing cardiac health program to make it culturally relevant and suitable for Indigenous Australians living in urban and remote communities.

Keywords:

Health Services; Indigenous; Heart Diseases.

Introduction

Cardiovascular disease (CVD, including stroke, heart and vascular diseases) is a major health problem for all Australians and is responsible for over a third of deaths annually [1,2]. In 2010, more than 50% of heart attack deaths were due to repeated events. CVD is more prevalent in Indigenous people, and they do not receive the same level of care for primary and secondary prevention, treatment and management as non-Indigenous Australians[3]. Only 3% of remote Indigenous patients were shown to be fully engaged in cardiac rehab [4]. Perhaps as a consequence, Indigenous people are 2.3 times more likely to die within four years of a cardiac event.

Cardiac rehab programs have been proven to be effective in preventing the recurrence of cardiac events and readmission to hospitals [5] but uptake by those eligible for cardiac rehab has been low. Under utilization of these traditional, hospital or centre-based cardiac rehab program services is mainly due to personal barriers for the patient (motivational, cultural and financial) or system-related barriers (referral, timing, travel and accessibility of programs) [5].

To address barriers to cardiac rehab, CSIRO and Metro North Hospital and Health Services (HHS) designed and developed the "Care Assessment Platform" (CAP). This program used smart phones and the internet to deliver cardiac rehab in the patient's home to align with their lifestyle (Figure 1). The CAP is the first clinically validated mobile health delivery of cardiac rehabilitation in Australia, tested in a randomized controlled trial at Metro North HHS in 2012 [5]. The CAP program was delivered in accordance with national guidelines to address all components of a cardiac rehabilitation program. Of the 120 participants, 60 (50%) were randomly assigned to traditional cardiac rehabilitation (TCR) and the remaining to the Care Assessment Platform – Cardiac Rehabilitation (CAP-CR). Compared to the TCR, the CAP-CR had significantly higher rates of all three primary outcomes [5] (Table 1). Of the participants who did not complete the program, competing life demands and logistics accounted for the majority of attrition in the TCR group, whereas change of circumstances was the primary cause of attrition in the CAP-CR group. Only three of these latter participants cited difficulty in using information technology tools as a contributing factor.

Table 1– Mobile delivery significantly improves primary outcomes over traditional cardiac rehabilitation care[5]

Program	CAP-CR (*n*=60)	TCR (*n*=60)
Uptake	80% (*n*=48)	62% (*n*=37)
Adherence	94% (*n*=45)	68% (*n*=25)
Completion	80% (*n*=48)	47% (*n*=28)
Attrition	20% (*n*=12)	53% (*n*=32)

Participants in both programs had significant improvements compared to baseline measurements in secondary outcomes including dietary intake, mental health and functional capacity. Improvements in quality of life scores were only seen for the CAP-CR group, and these were significant compared to baseline measurements and compared to the TCR group [5].

These findings are similar to a smaller US study [6], which showed that patients who complemented their cardiac rehab with a smartphone health and activity diary were less likely to be rehospitalised than those who undertook rehab alone.

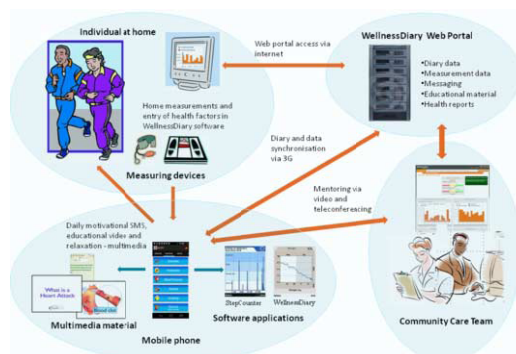


Figure 1 – The Care Assessment Platform for mobile health delivery of cardiac rehabilitation (source: M. Varnfield [7]).

At the time of writing, the authors are unaware of any phone-based cardiac rehab programs specifically for Indigenous communities, in Australia or internationally. Such programs may, however, be able to address the inequalities in provision of cardiac care due to flexible delivery of rehabilitation, health management and secondary prevention within the community.

The mobile cardiac rehabilitation health program has three main components:

- service delivery, encompassing the home care program, enrollment, the clinical portal and mentoring;
- education, comprising videos and motivational messages; and
- a smartphone app, containing a health diary and delivery of the educational material.

In this paper we describe changes to all three components to make the program suitable for rehabilitation of cardiac disease for Indigenous Australians. We have broadened the scope of the program to address risk factors for prevention in people with cardiac symptoms. Due to the popularity of iPhones, it was determined that the app needed to be available for both android and iOS phone users.

Methods

To customize the cardiac program for Indigenous Australians, involving Indigenous stakeholders was considered imperative. The research team followed National Health and Medical Research Council guidelines [8] by respecting culture and dignity, learning from Indigenous contributors, involving Aboriginal health workers, acknowledging the importance of family and carers, providing appropriate educational materials and including in the service delivery a setting comfortable for Indigenous people. In line with appropriate principles of engagement, CSIRO met with the Mayor and Council of a remote Aboriginal Community. During this meeting, the local Aboriginal Health Service received in principle agreement from the Council to provide a letter of support to hold discussions with the Health Service and members of the Community. Conversations were also held with a Indigenous health institute about the need to address cardiac health in the metropolitan community. The idea of conducting a pilot trial of the modified program was discussed with both communities.

Throughout the course of the program customizations, CSIRO actively engaged in meetings, workshops and interviews with six relevant healthcare providers who specialize in cardiac and/or Indigenous health, many of whose contributing staff members were themselves Indigenous. Weekly meetings were held with one group of stakeholders.

CSIRO identified an Indigenous production company to develop a culturally appropriate look and feel for the app using Indigenous artwork for the app icons. The company was provided with the original videos used in the urban program to be rescripted and remade to be culturally appropriate. An Aboriginal artist designed all the artwork for the videos and the app graphics.

Results

Customizations were made to all three main components of the mobile cardiac rehabilitation program: service delivery, education, and the app through which the program is delivered.

The service delivery model required minor modifications

Changes to the service delivery model included refinement of the rehabilitation program and inclusion of a focus on prevention, identification of relevant enrollment pathways, potential adaptations to the existing clinical portal, and mentor assignment and training. These are elaborated below.

The current rehabilitation program is conducted over a six week period with each week focusing on a specific set of themes supported by mentoring sessions over the phone. The themes designed by cardiac rehabilitation clinicians were not thought to require major changes, as they are clinical rather than cultural. Pending individual community feedback, the program has the flexibility to be extended to a twelve week program, covering the same material over a longer period, or implemented as a rolling program. For each week, theme-specific educational material would be provided to both rehabilitation and at risk patients, with weekly phone mentoring sessions in the rehabilitation program. At risk participants would have ad hoc access to a mentor as required.

Enrollment to the rehabilitation program would be through cardiac specialists based at proximal hospitals, with some recruitment through local Aboriginal health services. Enrollment for the 'at risk' program would be through identification of people with symptoms indicative of elevated risk of cardiac event determined by the local Aboriginal health service using a cardiac screening or decision support tool.

Generally, the format of the existing portal is considered appropriate, whereby following enrollment, patient data is entered into a website-based clinical portal. Throughout the program, data captured from patients are uploaded from the app to the clinical portal. A care delivery plan is tailored to each patient by the mentor with goals set according to their health condition profile and needs. Prior to each weekly phone call, mentors can review the data in the portal to inform their consultation. Based on feedback from an Aboriginal Health Unit, however, some minor modifications are required. These include building in the capacity to send compliance alerts when data points are missing, determining how often measurements need to be taken (eg. daily versus weekly) depending on the individual's cardiac health status, and including some positive measures, such as serves of fruit and vegetables each day.

It was determined that for some communities, it would be appropriate to establish a meeting place where participants meet regularly to discuss the program material and have their physiological measurements taken. The establishment of a meeting place was seen as a solution to several issues. Many communities have patchy mobile network coverage and community members are on prepaid mobile plan deals with limited data. The provision of free wifi to upload physiological measurements and download new app material would circumvent these issues.

To assist in developing community capacity and foster community ownership of the platform, mentors would be sourced from within the community wherever possible. Training packages include comprehensive guides for the mentoring role, the clinical portal, program content and the app. Where a meeting place is to be established, medical trainees from an appropriate institution would be sourced to take measurements and discuss educational material with participants under the supervision of a mentor.

The educational material required significant revamping

The original educational videos delivered as part of the six week rehabilitation program were caucasian-centric and heavily text based. Changes made to the video material

included reducing the length, simplifying the message and framing the videos in an appropriate context, with Indigenous artwork (Figure 2).

Figure 2 – The original text-based videos (left) were replaced with culturally relevant illustrated education delivery (right).

On each day of the six week program, a number of educational and motivational SMS messages are forwarded to the participant's smart phone. Due to issues with mobile coverage and data downloads in remote communities, these messages have been built into the app and appear as 'pop up' messages.

The current messages are considered to be too long, overcomplicated and lacking in persuasive delivery (Table 2). There is a strong feeling that the actual wording should be driven by individual communities. An example of the expected changes is provided in Table 2. The structure of the message delivery (week, day, time) aligns with the program themes and is therefore unlikely to undergo major change.

Table 2 – Examples of reworded motivational messages

Original message	Indigenous message
Did you know that there are driving restrictions after having a heart attack? It is important to discuss this with your doctor.	If your family depend on you to drive them around, then it's important to discuss this with your Doctor before getting behind the wheel after a heart attack because there are driving restrictions by Law.
Age is not a limitation to participation and success in Cardiac Rehabilitation. People of all ages do equally well!	You don't have to be young to participate and have success in Cardiac Rehabilitation. People of all ages do equally well!

The educational videos and messages are still a work in progress and will potentially undergo further changes through the course of community consultation.

The mobile app was customized for both iOS and android operating systems

The original app was functionally effective, but culturally inappropriate, for an Indigenous population, as it was text based, caucasian-centric and designed for individuals rather than community. An Indigenous production company was engaged to design new icons to improve the look and feel of the app. Based on feedback from Indigenous stakeholders (see Methods), the app icons needed to move away from the original text based format to a heavier reliance on visual appearance (Figure 3). They needed to be Aboriginal in context and aesthetically pleasing. The move to larger graphics was designed to accommodate for visual impairment, low literacy and persuasive appeal. All the artwork was designed by an Aboriginal artist.

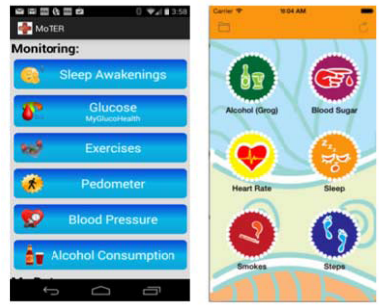


Figure 3 – The original text-based icons (left) were replaced with illustrated icons depicting Aboriginal artwork (right).

The heart symbol was chosen as the main program icon as it is intricately linked with family, health and life, and reflects the ultimate goal of the cardiac rehabilitation and prevention programs. Further customisations to the look and feel of the app were made to the sign-in screen, the screens in which measurements are recorded, the graphs and the general layout. These features were built using the artwork and colour schemes provided by the Indigenous production company.

Clear straightforward screen editors were developed with alternate entry options

A single screen editor was built for simple data entry, such as the count data; and individual screen editors for more complex measurements such as blood data (Figure 4). A generic screen editor was used for the remaining measurements.

Count measurements are provided for alcohol and cigarettes, with a custom option for the mentor to include, in consultation with the participant, any other substances recorded in the care plan (such as soft drinks). Pressing the icon takes the participant to recent history, where they can press the + sign to see their target for the day and update data using the + sign. Accidental data entry points can be deleted using the – sign. Alternatively, participants can press on the white box and enter a number directly (Figure 4). New functionality was added to count measurements whereby a single data point (ie. add one cigarette) can be entered by a long press on the respective icon on the home screen. This new functionality has been incorporated in the android version and is under development for the iOS version.

More complex screen editors are required for the collection of measurements relating to blood pressure and blood sugar. Data entry for blood pressure uses either a rolling counter or direct entry for systolic and diastolic values. While the android app can be updated directly from the blood pressure monitor using bluetooth, the iOS system requires manual input of values. Where blood sugar values are required as part of the care plan (ie. diabetes), values can be added as for count data. Once a value is entered participants nominate the context in which the blood sugar was measured (eg. before food).

For all other measurements, a generic screen editor automatically generates the editing panels for each measure.

Historical data can be viewed in list or graph form

For all health measurements, data can be viewed in list or graph format. Pressing any health icon displays the recent history page for that measure with a graph icon to display recent entries in graphical format. The graph displays data points for seven days (previous five days, today, tomorrow).

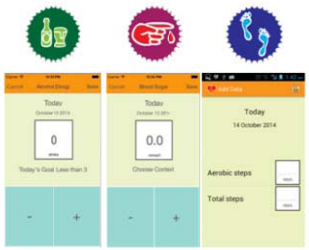


Figure 4 – Examples of simple, complex and generic screen editors built for iOS and android versions.

Educational videos can be viewed at any time

Prompts to access the educational material are meted out over the six weeks of the program. The five videos are now contained in the app and can be viewed at any time from the list function. As there are over 150 pop up messages, only the recent messages are now kept in the app. Motivational messages appear on the screen in the first instance in the same way text messages appear. Once read, the messages can be accessed from the list function icon, as for videos. Once the list display screen is full, older messages are deleted. New functionality is being developed to allow the text messages to have an audio option to accommodate for visual impairment, low literacy and sharing of message content in a group setting.

Functions outside the iOS–Android functionality overlap

App customizations were limited to functionality and appearances acceptable to both operating systems to maintain consistency. In the android version, however, there was the potential to include a traffic light system in the list function that was considered useful for the participant. This function allows a participant to see their program status at a glance. Green dots indicate that all health data has been entered, all scheduled videos have been viewed and that participation in the care plan is up to date. Yellow dots indicate that only some measures have been entered or videos still need to be viewed. Red dots indicate that no data has been entered and/or no scheduled videos viewed and/or that participation is required in the care plan. This functionality is under development in the android app. It is potentially possible to incorporate an equivalent system into the iOS app by adding compliance alerts to the threshold alerts contained in the clinical portal.

Summary of results

Modifications to the existing cardiac rehabilitation platform were made through initial community engagement, Indigenous health provider engagement, engagement with an Indigenous production company and in house software development. Modifications were made to all components of the platform (Table 3) including the service delivery model, the educational component, and the app based health diary. The revised app is available for both android and iOS operating systems.

Discussion

Closing the Gap between the health of Australia's Indigenous and non-Indigenous populations has been a campaign of the Australian Human Rights Commission since 2006 [9]. The improvement of Indigenous life expectancy, health and wellbeing was incorporated into health policy and formally supported by all Australian governments under the National Indigenous Reform Agreement which came into effect on 1 January 2009 [10]. In line with this agenda, this project aimed to develop a new model of service delivery by broadening an existing smartphone based cardiac rehabilitation program to encompass risk prevention, tailoring it for an Indigenous community and incorporating community capacity building into the implementation design. The aim of this model is to facilitate community autonomy of cardiac health management in the Indigenous population to assist in closing the gap in cardiac health outcomes.

The flexible delivery of mentored and supported cardiac rehabilitation can increase both uptake of, and adherence to, cardiac rehabilitation, as demonstrated in the CAP trial [5].

Table 3 – Summary of modifications

Component	Modifications
Service Delivery Model	
Rehabilitation and prevention programs	Programs developed and refined, identification of meeting place for discussions and taking of measurements.
Enrollment pathways	Aboriginal Health Services.
Clinical portal	Inclusion of positive measures.
Mentor assignment/ training packages	Mentors and trainees sourced from community. Training packages under development.
Educational Material	
Educational video scripts	New rescripted style frames and video, with Aboriginal artwork and emphasis on graphics.
Mentor messaging	Embedded messages. Rewording to make messages concise and culturally appropriate. Audio function for visual impairment, low literacy and sharing.
Look and Feel	
App program icon	Aboriginal artwork. Culturally relevant symbol.
Background graphic	Addition of a background using Aboriginal artwork.
Health diary	Aboriginal artwork, improved data entry functions, including one press entry. Larger graphics to accommodate visual impairment and low literacy levels and to add persuasive appeal. Traffic light code for missing data.
The Built Apps	App now available for both android and iOS operating systems.

The foundation of delivering a health program in a patient's home environment, rather than have them travel to a health setting, resulted in effective adherence to all components of the cardiac rehabilitation program. CAP-CR group adherence to attend to educational components and health measurements (≥80%) demonstrated patients' ownership of, and motivation to understand and address, their lifestyle and biomedical risk factors. The appeal of the home delivery was supported by the CAP-CR group showing reduced anxiety levels over those in the TCR group [5]. Key lessons from the trial can now be applied to a pilot of this program in an Indigenous community.

The introduction of a technology based program to provide information, mentoring and monitoring for a targeted intervention in Indigenous communities is likely to be well received. Indigenous people are rapid adopters of smart phones with a surprisingly high uptake in communities where there is coverage [11,12], hence smart phones provide an ideal platform for health engagement. While numerous succesful smartphone health apps are currently available on the global

market [13], few have been specifically designed for Indigenous populations. In Australia however, the recent development of smartphone apps by and for Aboriginal groups [14,15], suggests this technology will be key for healthy Indigenous futures.

Given the adoption of mobile phone use in Indigenous communities, similar adherence to educational delivery is likely, if the material and content modifications are suitable for an Indigenous audience. Similarly, adaptations to the health monitoring and program delivery through a culturally appropriate smartphone app, together with engagement with their peers and health service support, are likely to result in similar improvements in key health outcomes as those seen in the CAP-CR trial, including physical activity, psychological wellbeing (anxiety and depression), and overall quality of life.

Health engagement should be combined with capacity building to provide communities with the resources and knowledge to identify, prevent and manage cardiac risks within the community. In this program, community-sourced mentors and trainees would contribute to community capacity. Central to maintaining this capacity is the establishment of a gathering place for participants to meet regularly to discuss the program material and have their physiological measurements taken. Trainees play a crucial social role as well as ensuring consistency in measurements and hygienic care of equipment at the meeting place. Anecdotes from exemplars of community based health education programs [16] suggest that greater understanding and awareness is derived from group participation. Possibly of greatest import in a community environment, the meeting place fosters a social network system for participants and facilitates a whole of community approach to healthcare.

Integrated service delivery model for Community

The service delivery model for an Indigenous community would involve integrating the cardiac health program into the existing community health infrastructure (Figure 5).

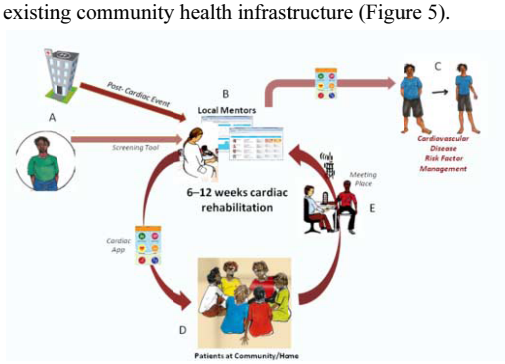


Figure 5 – Enrollment to the program would be through hospital-based cardiac specialists or screening tools (A). Mentors sourced from within the community would use the clinical portal to develop a care plan (B). People at risk (C) and those rehabilitating (D) can make and track data entries of their health and wellbeing using the cardiac app on their smart phone. Participants meet regularly in a meeting place to discuss the program material and have their physiological measurements taken (E).

Conclusion

This is the first smartphone app delivered cardiac health program designed specifically for Indigenous Australians. This novel program is now ready for further community consultation and pilot trial in an Indigenous community.

Acknowledgements

For their contribution to the development of the Indigenous cardiac program, CSIRO would like to thank: Gurriny Yealamucka Health Service, Yarrabah Aboriginal Community, Institute for Urban Indigenous Health, Carbon Media, PWC Indigenous Engagement Unit, Queensland Health Aboriginal and Torres Strait Islander Health Unit, Queensland Health Tele-health Unit and the Queensland Health Indigenous Cardiac Outreach Program.

References

[1] Australian Institute of Health and Welfare (AIHW) 2010. Australia's Health 2010. Australia's Health no. 12. Cat. no. AUS 122. Canberra: AIHW.

[2] AIHW 2014. Australia's Health 2014. Australia's Health series no. 14. Cat. no. AUS 178. Canberra: AIHW.

[3] AIHW 2011. The Health and Welfare of Australia's Aboriginal and Torres Strait Islander People, An Overview 2011. Catalogue No.: AIHW 42. Canberra: AIHW.

[4] Shepherd, F., Battye, K., & Chalmers, E. 2003. Improving access to cardiac rehabilitation for remote Indigenous clients. Aust N Z J Public Health, 27(6), 632-636.

[5] Varnfield M, et al. 2014. Smartphone-based home care model improved use of cardiac rehabilitation in post-myocardial infarction patients: results from a randomised controlled trial. Heart doi:10.1136/heartjnl-2014-305783.

[6] Widmer, R.J., et al. 2014 The augmentation of usual cardiac rehabilitation with an online and smartphone-based program improves cardiovascular risk factors and reduces rehospitalizations. J Am Coll Cardiol. 63(12_S).

[7] Varnfield, M. 2014. PhD Thesis: Evaluation of Information Communication Technology platforms to Improve Self-management of Chronic Disease.

[8] NH&MRC 2005. Strengthening cardiac rehabilitation and secondary prevention for Aboriginal and Torres Strait Islander peoples. Australian Government.

[9] Australian Human Rights Commission 2005. Social Justice Report 2005: The Indigenous Health Challenge.

[10]COAG 2008. National Indigenous Reform Agreement (Closing the Gap). Council of Australian Governments.

[11]IRCA 2011. IRCA Submission to the Regional Telecommunications Enquiry 2011. Indigenous Remote Communications Association.

[12]Department of Industry and Resources 2006. Response to the DCITA Discussion Paper on Mobile Connect: Improving Regional Mobile Phone Coverage. Perth: Government of Western Australia.

[13]Sutton, M. and Fraser, M. 2013. The rise of smartphone health and medical apps. Medical Devices, Australian Life Scientist, January.

[14]http://www.naccho.org.au/naccho-app/

[15]Ridani, R., Shand, F., Tighe, J., Christensen, H. 2013. An app for suicide prevention among Indigenous Youth. Black Dog Institute, UNSW and Men's Outreach Service.

[16]IUIH 2012. Submission Report. Inquiry into the contribution of sport to Indigenous wellbeing and mentoring. Institute for Urban Indigenous Health.

Address for correspondence

David Hansen, Australian eHealth Research Centre, Royal Brisbane and Women's Hospital, Herston Queensland, 4029, Australia.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-348

A Serious Game for Upper Limb Stroke Rehabilitation Using Biofeedback and Mirror-Neurons Based Training

Diego João Cargnin[a], Marcos Cordeiro d'Ornellas[a], Ana Lúcia Cervi Prado[b]

[a] *Laboratório de Computação Aplicada, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*
[b] *Hospital Universitário de Santa Maria, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*

Abstract

Upper limb stroke rehabilitation requires early, intensive and repetitive practice to be effective. Consequently, it is often difficult to keep patients committed to their rehabilitation regimen. In addition to direct measures of rehabilitation achievable through targeted assessments, other factors can indirectly lead to rehabilitation. Current levels of integration between commodity graphics software, hardware, and body-tracking devices have provided a reliable tool to build what are referred to as serious games, focusing on the rehabilitation paradigm. More specifically, serious games can captivate and engage players for a specific purpose such as developing new knowledge or skills. This paper discusses a serious game application with a focus on upper limb rehabilitation in patients with hemiplegia or hemiparesis. The game makes use of biofeedback and mirror-neurons to enhance the patient's engagement. Results from the application of a quantitative self-report instrument to assess in-game engagement suggest that the serious game is a viable instructional approach rather than an entertaining novelty and, furthermore, demonstrates the future potential for dual action therapy-focused games.

Keywords:

Computer Games; Patient Engagement; Rehabilitation; Stroke.

Introduction

Cerebrovascular diseases have a great impact on an individual's health. These diseases are one of the leading cause of hospitalizations, mortality and disability in the world population, surpassing heart disease and cancer, the two leading causes of death in industrialized countries. A stroke is defined as a sudden neurological impairment due to vascular injury, including lesions caused by hemodynamic disturbances. The severity of the condition and the functional impairment varies according to the affected area and to the vascular structures, and may cause cognitive, sensory and or functional deficit.

The advancement of technology has allowed the development of procedures enabling exploration of brain activity during rehabilitation therapies, where brain activity is a vital factor in motor rehabilitation. Recently, in order to determine the most effective and efficient techniques, a branch of study has focused on trying different methodologies relating biofeedback and mirror-neurons.

The treatment of neuromuscular disorders including hemiplegia and hemiparesis requires training of the remaining functional capacity. Thus, the motor training is based primarily on a rehabilitation process, in which the feedback is an essential component. Biofeedback goals in rehabilitation include the optimization of muscle contraction and elimination of mass movements and synkinesis. Therefore, patients can improve the symmetry and synchronization between the affected side and the healthy one, seeking functionality and spontaneity of movement. Mirror neurons are activated both when performing an action and during the observation of the same action performed by another person. As they seem to integrate observation and action, mirror neurons have been the focus of studies on how humans understand others and to what extent they are able to share experiences. Such integration includes an internal representation involving the same neural structures related in the execution of the observed action. It also has been considered as a key for rehabilitation purposes.

A Natural User Interface (NUI) represents a current concept that deals with the interaction between human beings and electronic devices. Through the identification of gestures, voice commands, expressions, body movements or detection of human body parts such as face, hand or joints, new games have been built, incorporating this technology.

The entertainment factor aside, serious games are designed to solve problems in several areas. Considering the health application, one can highlight the important contribution of these games to physical therapy and the rehabilitation of patients with motor disabilities secondary to neurological diseases. These patients require physical therapy programs appropriate to their problems in order to improve their quality of life, as much as possible. However, traditional physiotherapy is often boring and repetitive and can discourage patients, causing them to abandon their rehabilitation programs before they have finished.

This paper presents a serious game application for upper limb stroke rehabilitation based on biofeedback and mirror-neurons-based training, focusing on patients with mild or moderate stroke incapacity. The game helps the therapist by assessing the quality and efficiency of the patient's treatment. The patient participates by playing a game that uses quantitative measures to assess in-game engagement while incorporating existing interfaces for human-computer interaction.

The paper sections are laid out in the following order: Related Work, Hemiplegia and Rehabilitation, Game Framework and Design Elements, In-Game Engagement Assessment, Results, Discussion, and Conclusions and Further Work. The Related Work section displays some relevant information on the approaches and technologies concerning the rehabilitation process and the application of serious games within the field of rehabilitation. The section titled Hemiplegia and Rehabilitation briefly illustrates the causes, symptoms, and treatment of hemiplegia. The Game Framework and Design Elements section discusses the necessary components for serious game development, primarily focusing on rehabilitation. The following section reports initial upper extremity assessment based on a valid control group of physical therapists. The Results, Discussion, and Conclusion and Further Work sections present a

general discussion on game design, engagement, and use in upper limb rehabilitation, including therapist perspectives.

Related Work

Current research indicates that in the target audience, visibility and feedback may be considered for the development of serious games for rehabilitation, since they are important human factors [1]. The principles of game design that have particular relevance to rehabilitation are meaningful play and challenge. The former is the relationship between player's interactions and system reaction, while the latter aims at maintaining an optimal difficulty in order to engage the player [2][3]. Some important criteria for the design and evaluation of serious games in the rehabilitation area have been identified as game interaction and interface, number of players, genre, adaptability, performance feedback, game monitoring and portability [4]. It was also observed in [5] that serious games must ensure that patients are correctly performing the exercises as well as provide a motivating atmosphere for therapy, so as to yield maximum impact on the rehabilitation process.

Several technologies are used in serious games for physical therapy and rehabilitation, namely virtual reality and sensor networks. Games for physical therapy with severe injuries in the upper limb are proposed in [6]. The neurogame therapy, used for the rehabilitation of patients with severely compromised movements after stroke, is depicted in [7]. To improve the communication between physical therapists and patients with disabilities, a set of requirements for an authoring tool to enable therapists to create exercise animations for home viewing is presented in [8].

The benefit of almost all of these technologies is that they are inexpensive and can be implemented in any clinical environment. Perhaps their greatest benefit is that they provide real time feedback to the player and if designed effectively, they are more enjoyable than traditional exercises, and may therefore increase adherence.

Hemiplegia and Rehabilitation

Cerebral Vascular Accident

Stroke is the second major cause of death worldwide and is one of the greatest causes of overall disease burden. Each year more than 5.5 million deaths are attributed to cerebral vascular accidents. Despite this, less than one-third of countries take minor measures against it such as mortality reporting [9]. In 2010, 31% of stroke victims were children (aged <20 years old) and adults between 20 and 64 years old, with a higher overall proportion of strokes noted in the elderly [10]. In addition, [16] presents a complete statistical study on stroke by country income which reveals that fatality and incapacity rates are higher in low and middle-income countries. The high death rates among older people (aged > 75) indicates that younger adults suffer most from disability-adjusted life-years (DALYs).

Hemiplegia and Hemiparesis

Hemiplegia is a condition of full or partial paralysis of one side of the body that can occur secondary to a brain tumor, vascular accident, stroke, childhood head trauma, or congenital or perinatal injury. Unilateral impairment of hand and arm function are the most disabling symptoms of hemiplegia. Motor and sensory impairments in people with hemiplegia may compromise movement efficiency [11]. Hemiplegia is similar to hemiparesis, but it is far more serious. Hemiparesis is usually characterized by one side of the body being affected not by paralysis, but a less severe state of weakness.

Serious games are one possible clinical intervention for people living with the effects of stroke. In a serious game, a computer generates an engaging environment in which the patient must perform goal-oriented tasks that they find difficult. In this instance a serious game experience is similar to a regular video game experience, the key difference being the main goal: serious games aim to elicit physical movements. This treatment can be effective as the brain is a resilient organ that produces new neural connections in response to external stimuli and activity to compensate for damage.

Rehabilitation of Individuals with Disabilities

Physical therapy is the leading therapy in the treatment of individuals with disabilities. Therapy goals can include enabling people to reach and maintain optimal physical, sensory, psychological, social, and intellectual levels. Rehabilitation provides disabled people with the necessary tools to achieve independence and self-determination. To achieve these goals, physical therapists may apply various techniques to improve physical and mental capacity of the impaired patient.

In addition to sensory or motor deficit, people with disabilities can develop secondary pathological conditions such as shoulder pain, which affects 34% to 85% of those involved in the treatment. [12]. Rehabilitation and recovery constitutes a great challenge for both the complexity of the functions lost and the high incidence of shoulder pain, resulting in a negative impact on the overall rehabilitation process.

Game Framework and Design Elements

Natural User Interfaces in Physical Therapy

Nowadays, NUI is highly used in rehabilitation, generating research and applications that contribute to patient performance and facilitating the assessment of patient progress. NUI-based physical therapy has a functional, concrete, and challenging context, bringing a direct benefit for both the people with disabilities and the physical therapist through the adaptability of these systems. Advanced technologies are used to produce simulated interactive and multidimensional environments. Visual devices and hardware devices for body tracking are important to immerse patients in a virtual environment and give them the ability to modify the environment based on the rehabilitation goals.

NUI can also be considered more fun and appealing than the traditional set of exercises. The use of new tools, which certainly increases the quality of service they provide, also allow greater coverage without increasing risks. By using a NUI console, patients are encouraged to play games and stay engaged, rather than becoming bored. The console offers a variety of opportunities to work on physical abilities, depending on which game is selected.

Biofeedback and Mirror-Neurons

Various rehabilitation techniques such as biofeedback training to stimulate motor functions, have been studied and introduced in regular therapeutic techniques. Biofeedback techniques allow the patient to have an immediate response to exercise in view of the effects caused by neuronal stimulation, demonstrated with the aid of devices and tests that give a noticeable response externally. Thus, patients can monitor their motor and cognitive improvement, even if not visible at first sight, which reinforces the continued treatment while

stimulating other areas of the brain to recognize biofeedback response.

The biomechanical biofeedback involves measuring the movement, posture and control forces produced by the body [13]. Microsoft Kinect may represent a reasonable alternative for obtaining data that is relevant to the analysis due to its ability to map the body's joints of the patient participating in the game. The camera sensor helps physical therapists to examine movement quantitatively, which is beyond the game learning aspects, working as a learning speed-up tool so that the patient can quickly grasp the meaning. Based on these aspects, it is possible to use the Kinect to analyze the optimal position of upper limbs through the joints to make measurements of abduction and adduction as well as extension or flexion.

A small window in the game interface shows relevant information through a biofeedback character in stick figure (i.e. a digital goniometer), whereby patients as well as physical therapists can monitor real time movements. Based on this information, physical therapists can correct possible errors (i.e., movement or postural deviation errors) which can occur during the game. Physical therapists are allowed to monitor hardware and software setup, tailoring the game according to the levels of the treatment. The interface can be programmed to show several variables that may help rehabilitation and recovery.

A unique property of mirror-neurons is the ability to act both during the implementation of the action as well as in the observation of specific actions. This property allows humans to perceive the intentions of others through action observation. Therefore, observation is essential in imitation-based learning and empathy. While working with upper limb stroke rehabilitation, it has been observed that the mirror-neuron system in adults is still plastic and artificial activation of this system can provide the basis for cerebral post-stroke rehabilitation patients. The use of active observation as a form of rehabilitation is an asset in patients with severe paresis, for whom rehabilitation can be difficult [14]. Imitation-based learning requires a complex cognitive function that is constructed step by step, including motor observation. Therapies that activate mirror-neurons may be used along with the treatment, seeking a better post-stroke rehabiliation. The proposed serious game includes video recording of motor activities.

Playability and Usability

There are a number of aspects in the usability of digital games where the factor efficiency and effectiveness should be secondary to the satisfaction factor. Thus, game engagement must be addressed on a different perspective of the working software. The main point is to identify where the challenge must be and keep any other aspect out of this context according to the traditional recommendations of usability working software. Thus, the designer should create challenges that are part of the fun of achieving the goals of the game rather than undermine the patient's experience and effectiveness in tasks that should require some effort.

Serious Game into Play

After analyzing regular exercises with physical therapists in a rehabilitation ward, a decision was made in favor of shoulder abduction, as it was identified as the most common issue across various neurological diseases. Shoulder abduction is important for upper limb functions, as more distal functions depend on a stable shoulder.

Game prototyping is a high quality product development method in which a prototype is built, tested, and then reworked as necessary until an acceptable prototype is achieved. As a preliminary step in the prototype evolution, the target audience was defined. The proposed serious game is intended to be a complementary treatment to people with upper limb impairments after stroke in a mild or moderate condition.

The analysis of existing video games on the commercial market was important to support the concept and the choice for the game engine that integrates seamlessly with the Kinect. Later, it evolved to a concept that was more focused on patient engagement through a concept and argument that would provide an immersive environment necessary for their motivation. The biofeedback integration approach with mirror neurons served as support for the development of the serious game for physical therapy rehabilitation.

The game was implemented using Unity3D game engine and Kinect. Unity3D is a commercial engine, providing the developers with a large range of useful tools and scripts, as well as the underlying graphical tasks for game development. Even though such an engine is commonly used for entertainment games, a systematic move into serious game development has been recently detected. Given its integrated development environment with some set of game reusable components, Unity3D, in close cooperation with Kinect software development kit (SDK), reduces the workload needed to produce the game, allowing developers to concentrate on the higher levels modules of the game.

A screenshot of the serious game for upper limb stroke rehabilitation is shown in Figure 1. Control panels on the left and right side are intended for software setup. The goal of the game is to harvest as many fruits from the tree as possible and put them in the baskets within a certain time period. The winner will be the one who harvests more fruit. The angles shown in the interface are the measures of player abduction. Calculations are made based on the position of the joints. The shoulder angle is measured by the elbow, shoulder and pelvic tilt while the player is measured by the bowl and neck to the ground.

Figure 1 – Screenshot of the serious game for upper-limb stroke rehabilitation

The player collects the fruits by simply moving his/her hands until it touches the object on the screen. A Unity3D hand collider performs the detection so fruit coordinates become hand coordinates. Once the fruit is following the character's hand, the player must bring them to the basket where another collider transfers the coordinates to the fruit basket.

In-Game Biofeedback and Rapid Prototyping

After the game scenarios have been built, a feedback from the control group of therapists is needed as a particular step in the rapid game prototyping approach. The control group watches a therapy session of players and afterwards discusses it. This ensures that the scenarios are well adapted for the therapy and

appealing to the audience. After this step, strategies, game design ideas and concerns are evaluated and considered for the next prototype.

The effective user-centered design (UCD) issues on prototyping recommend prevention of implementing the final solution directly. Therefore, the serious game proposed is developed in an iterative manner. After the first prototype, UCD continues in an iterative process of evaluation and redesign. Observations and recommendations on the serious game are made by the control group of therapists while performing typical tasks for the game experience both in the process of doing the exercises and their relevant opinions about the game.

In-Game Engagement Assessment

Recruitment

Participants were instructed to play several rounds of the serious game. During the prototype development based on user-centered design, the game metaphor was chosen because it can be easily used to evaluate selection and reaching performance. Thus, the serious game is mildly challenging, has a player motivating potential, is not gender-specific and can be played by people of all ages. Furthermore, the game is easy to learn and understand.

Forty-eight participants (36 female and 12 male) took part in the in-game engagement assessment. Two participants (1 female and 1 male) had to be excluded due to impaired vision and the not meeting eligibility criteria. The age of participants ranged from 20 to 43 years. Participants had no prior knowledge of the experiment. They were recruited via academic and technical staff from the neuro-rehabilitation ward, including physical therapy undergraduates and graduate students, physical therapists and lecturers from the physical therapy department. Participants in the control group were required to have normal or corrected to normal vision as well as no physical impairment, to be eligible.

Preliminary Procedures and System Setup

Experiments were conducted on weekdays from 10 AM until 6 PM. The average time for an experimental session was 20 minutes, including setup and cleanup. Upon arrival, participants were informed about the nature of the experiment, which was the usability evaluation rather than a therapeutic application. Participants were also informed of the rehabilitation goals of the game, anonymity and ethical approval and the experimental procedure.

After description of the procedure, each participant signed an image consent release form and an informed consent form. Shoulder, elbow, and wrist joints were tracked and mapped to the avatar in the application scenario for upper limb rehabilitation. Since hemiplegia is characterized by unilateral impairments, the game was developed in such a way that the player could use both sides of the body.

Participants were led to the center of the room for the setup and sensor calibration. After this step, participants played the game levels in accordance with the goals set by the physical therapist. Each game session was set to 15 min, but participants could finish early. After the experiment, participants filled out a digital short-form version of the Game Experience Questionnaire (GEQ) to assess in-game engagement. Participants did not receive any kind of reward for their valuable participation in the evaluation experiment.

Measuring Serious Game Experience

A short version of GEQ was used to assess in-game assessment. The questionnaire has been validated as an effective tool both for commercial and non-commercial games. In this paper, only the first fourteen items were considered as a general metric to evaluate game experience. The findings cannot be considered to evaluate technical details in the user-centered development process.

The questionnaire links several game-related measures and implemented focusing on group research, consisting of seven dimensions: sensory and imaginative immersion, flow, competence, tension, challenge, positive and negative affect. Each dimension was measured by two questionnaire items on a Likert scale (0: no agreement with the statement to 4: complete agreement with the statement). The questionnaire[1] was clinician-administered upon participants' completion of the experiment.

Results

Results obtained from GEQ categories are represented in Figure 2. Our serious game prototype for upper limb stroke rehabilitation performed very well, considering the control group of therapists, in the positive affect dimension (M 3.22, SD 0.60). Players had a lot of fun playing the game and felt happy. Significant feelings of competence (M 2.20 SD 0.34) and challenge (M 2.8 SD 0.53) were observed. Participants lost their sense of time quite early in the game, suggesting substantial flow (M 2.98 SD 0.62).

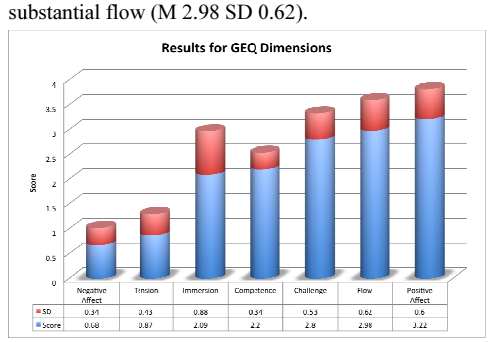


Figure 2 – Results of In-Game Engagement Assessment

The game performed reasonably well in the immersion dimension (M 2.09 SD 0.88). Almost all participants were visually delighted by the interface prototype (3.2) and were moderately impressed (2.04). The majority of players found that the game was neither tiresome nor boring (M 0.68 SD 0.34), suggesting that the prototype performed well in the negative affect dimension. A small amount of tension was observed (M 0.87 SD 0.43).

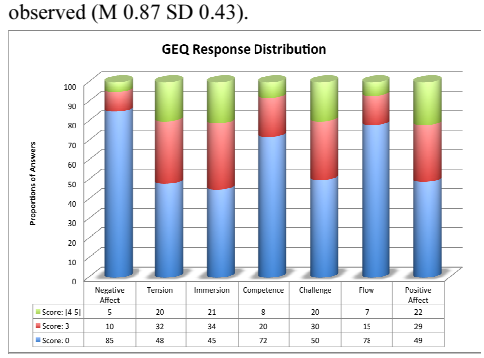


Figure 3 – GEQ Response Distribution

[1] The Game Experience Questionnaire captures game experience and playing experience based on a number of items. See more in http://www.allaboutux.org/game-experience-questionnaire-geq

Figure 3 presents the GEQ response distribution. T-tests were conducted to determine which response distributions differed significantly from the no agreement with the statement (score 0). A significant number of responses were either rather or exceptionally positive for items ($p < 0.05$) for dimensions competence $t_{14} = 2.92$, tension $t_{14} = 2.84$, positive affect $t_{14}= 2.41$, and sensory and imaginative immersion $t_{14}= 2.23$. Participants scored 0 exhaustively for negative affect, flow, and challenge dimensions.

Discussion

Results achieved during the in-game assessment suggest that the serious game prototype performed quite well in terms of usability and game experience. The serious game for upper limb stroke rehabilitation is rather new in the neurorehabilitation ward of the HUSM. Nevertheless, results gathered from the in-game assessment can be used as a standing point for new prototypes derived from the same game metaphor. Limitations include use of the short version of the GEQ scoring system and the need of functional evaluation using Capabilities of Upper Extremities Questionnaire (CUE-Q)[15].

Conclusion and Further Work

The main contribution of this work was the development of a serious game for upper limb stroke rehabilitation based on NUI, which allows the achievement of customized physical therapy sessions in real time. Measurements taken by the therapist allow the exercises to be kept within the limits of each individual, acting as an interface to the choice of how much and how long each movement should be performed based on competition and the entertainment stimuli provided by the game.

Besides assisting in the treatment of post-stroke rehabilitation, the serious game allows patients to enjoy the game and encourage movements that are generally repetitive. This extra motivation accelerates recovery and can be achieved by biofeedback and mirror-neurons through natural user interfaces. Results obtained during the in-game assessment were relevant to the evaluation of the prototype.

As future work, the interface may be enhanced using animation techniques to achieve a more engaging experience. A psychological and functional evaluation is planned in the near future to support therapist decisions over the course of treatment.

Acknowledgments

This research was partially supported by CAPES through the PROCAD Program. The participation of all individuals involved in the testing and evaluation step is thankfully acknowledged.

References

[1] Laver K, George S, Thomas S, Deutsch JE, and Crotty M. Virtual Reality for Stroke Rehabilitation. Eur J Phys Rehabil Med. 2012 Sep;48(3):523-30.

[2] Deterding S, Dixon D, Khaled R, and Nacke L. From Game Design Elements to Gamefulness: Defining "Gamification." In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (New York, NY, USA, 2011), 9–15.

[3] Holman C, Aguilar S, and Fishman BJ. GradeCraft: What Can We Learn From a Game-inspired Learning Management System? In: Proceedings of the Third International Conference on Learning Analytics and Knowledge (2013), 260–264.

[4] Wouters, P, Nimwegen C, Oostendorp H, and Spek E. A Meta-analysis of the Cognitive and Motivational Effects of Serious Games. Journal of Educational Psychology, Vol 105(2), May 2013, 249-265.

[5] Delbressine F, Timmermans A, Beursgens L, Jong M, Dam A, Verweij D et al. Motivating Arm-hand Use for Stroke Patients by Serious Games. In: Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, p. 3564–3567, 2012.

[6] Carter AR, Connor LT, and Dromerick AW. Rehabilitation After Stroke: Current State of the Science. Current Neurology and Neuroscience Reports. 2010;10(3):158–166.

[7] Marston HR, and Smith ST. Interactive Videogame Technologies to Support Independence in the Elderly: A Narrative Review. Games for Health Journal, 1(2): 139-152, 2012.

[8] Rahman MA, Qamar AM, Ahmed MA, and Basalamah S. Multimedia Interactive Therapy Environment for Children Having Physical Disabilities. In: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (ICMR '13. ACM, New York, NY, USA, 313-314, 2013.

[9] Mukherjee, D., & Patil, C. G. (2011). Epidemiology and the global burden of stroke. World Neurosurgery, 76(6 SUPPL.), 85–90.

[10] Krishnamurthi, R. V, Feigin, V. L., Forouzanfar, M. H., Mensah, G. a, Connor, M., Bennett, D. a, Foundation, M. G. (2013). Global and regional burden of fi rst-ever ischaemic and haemorrhagic stroke during 1990 – 2010 : fi ndings from the Global Burden of Disease Study 2010. The Lancet Global Health, 1(5), e259–e281.

[11] Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ et al. Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline. Stroke. 2005;36:e100-43.

[12] Golomb MR, McDonald BC, Warden SJ, Yonkman J, Saykin AJ et al. In-Home Virtual Reality Videogame Telerehabilitation in Adolescents With Hemiplegic Cerebral Palsy. Archives of Physical Medicine and Rehabilitation 91, no. 1 (1, 2010): 1-8.e1.

[13] Oonagh OM, Persson, UM, and Caulfield B. Biofeedback in Rehabilitation. Journal of NeuroEngineering and Rehabilitation, 10:60, 2013.

[14] Buccino G, Solodkin A, Small SL. Functions of the Mirror Neuron System: Implications for Neurorehabilitation. Cognitive and Behavioral Neurology, Vol. 19(1): 55-63, 2006.

[15] Oleson CV, and Marino RJ. Responsiveness and Concurrent Validity of the Revised Capabilities of Upper Extremity-Questionnaire (CUE-Q) in Patients With Acute Tetraplegia. Spinal Cord 52, 625-628 (August 2014.)

[16] Johnston, S. C., Mendis, S., & Mathers, C. D. (2009). Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. The Lancet Neurology, 8(4), 345–354.

Address for correspondence

Diego João Cargnin
diegocarnin@gmail.com
Av. Roraima, 1000
Prédio 7 – Anexo B- Sala 388
UFSM - Campus Universitário
97105-900 Santa Maria-RS - Brasil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-353

Usability and Safety of Software Medical Devices: Need for Multidisciplinary Expertise to Apply the IEC 62366: 2007

Sabrina Bras Da Costa[a,b], M.C. Beuscart-Zéphir[a], J-M. Christian Bastien[b], Sylvia Pelayo[a]

[a] *INSERM CIC-IT 1403 Evalab, CHU Lille, UDSL EA 2694, Lille University, F-59000, Lille, France*
[b] *PErSEUs (EA 7312), Univ Lorraine, Metz, France*

Abstract

Software medical devices must now comply with the "ergonomics" essential requirement of the Medical Device Directive. However, the usability standard aiming to guide the manufacturers is very difficult to understand and apply. Relying on a triangulation of methods, this study aims to highlight the need to combine various expertises to be able to grasp the standard. To identify the areas of expertise on which the usability standard relies, an analytical review of this document was performed as well as an analysis of a discussion forum dedicated to it and an analysis of a case study of its application for CE marking. The results show that the IEC 62366 is a usability standard structured as a risk management one. It obviously requires Human Factors/Ergonomics expertise to be able to correctly identify and prevent risks of use errors, but it also requires risk management expertise to be able to grasp the issues of the risk analysis and master the related methods.

Keywords:
Usability; Medical Device Software; Use Error; Risk Management; Standardization.

Introduction

It is widely accepted now that software can be either a component of a medical device (MD) or a medical device [1]. Thus, software for certain purposes is subject to the same regulation as MDs. Since 2010, software MDs must also comply with the "ergonomics" essential requirement aiming to ensure patient and users health and safety by preventing risks of use errors. The manufacturers must now integrate a Usability Engineering Process (UEP) in their MD design and development cycle and document it for CE marking. To comply with this requirement, the IEC 62366 standard [2] has been harmonized with the MD Directive to guide manufacturers.

However, this usability standard proves to be very difficult to be understood and applied by manufacturers, but also by competent authorities or even usability experts [3]. First, the IEC 62366 suffers the same design flaws as most of the standards [4]: references to many other standards, specific terminology, too general descriptions, etc. Futhermore, this standard is at the intersection of several areas of expertise: (i) it is a usability standard relying on a substantial Human Factors/Ergonomics (HF/E) conceptual and methodological expertise and (ii) it specifies a UEP that must be integrated into two other complex processes, the risk management process (RMP) [5] and the quality management process (QMP) [6], each one needing a specific expertise to be mastered.

Relying on a triangulation of methods, this study aims to highlight the need to combine these various areas of expertise to be able to correctly understand and apply the IEC 62366 usability standard. First, an analytical review of the structure and the content of the IEC 62366 standard document was performed to identify the domains of expertise on which it relies. Then, an analysis of a discussion forum dedicated to the IEC 62366 was undertaken to identify the difficulties expressed by the participants and the specific areas of expertise to which these difficulties were related. Finally, we took the opportunity to observe the real expertise-related difficulties for HF/E experts and a manufacturer during a case study of the application of the standard for a MD CE marking.

Methods

Analytical review of the standard document

A detailed analysis of the document was performed; only the results illustrating the purpose are presented here.

References

All the references made by the IEC 62366 to standard documents have been identified and linked to the area of expertise to which they refer for (i) the normative parts of the IEC 62366 (Articles 1 to 7) and (ii) the Bibliography section.

Terms and definitions

For the “Article 3. Terms and definitions”, the analysis goes further with a particular comparison. The terms defined without specific reference, i.e. exclusively decreed by the IEC 62366, were compared with classic HF/E terms defined in the ISO 9241 standards. This series of essential standards in the HF/E field is a reference in the domain and includes a wealth of information that covers every aspect of usability.

Usability Engineering Process

The Article 5 "Usability Engineering Process" describes specifically the requirements for the UEP implementation as well as its documentation. Its analysis also goes further than the analysis of the references with a comparison of the description of the processes between the IEC 62366 and the admitted standards of the different areas of expertise: (i) the ISO 13407:1999 [7] and the ANSI/AAMI HE74:2001 [8] for the UEP process, (ii) the ISO 14971:2007 for the risk management process, and (iii) the ISO 13485:2003 for the quality management process. The comparison was focused on

the objectives of each standard, the required steps and methods for each process and the intended level of requirements (normative or informative requirement).

Analysis of the discussion forum

Data selection and extraction

The queries "*IEC 62366 forum*" and "*IEC 62366 blog*" were performed in the Google® search engine. The first result was a discussion forum named Elsmar Cove with multiple threads on the IEC 62366. In December 2012, the query "62366" was performed in this forum's search engine. All the threads discussing the IEC 62366 standard were included in the study; those that only cited the standard were excluded.

Data analysis

An automatic textual data analysis of the threads' content was performed with the ALCESTE software [10]. It aims to extract the strongest significant structures of the text, named the "lexical classes", so as to draw the essential information contained in the textual data. After conducting a clean to avoid methodological biases, the ALCESTE software performed the analysis:

i. a basic vocabulary dictionary was created: lexical forms were simplified to gather together forms with the same lexical roots (e.g. "*required*", "*requirement*" or "*requirements*" were turned into "require+");

ii. the corpus was divided into Elementary Context Units (ECU) which correspond to the "units of text" within which ALCESTE was able to calculate words' co-occurence frequencies;

iii. ALCESTE crossed the ECU and the presence/absence of the lexical forms to form classes. The Chi-Square Test revealed the associative strength between a word and a class. For a class, ALCESTE was also able to compute a list of words that were characteristics of the class.

Application of the standard: a case study analysis

Context of the study

MDoloris Medical Systems® had designed an innovative analgesia monitor named PhysioDoloris® (Figure 1). It provided a new pain indicator called A.N.I. (Analgesia Nociception Index) to support better management of the monitoring of a patient's pain during general anesthesia. HF/E experts were asked to (i) perform the usability verification and validation as required by the IEC 62366 standard and (ii) support the UEP documentation for CE marking. The manufacturer and the HF/E specialists had never applied the standard at the time of the study.

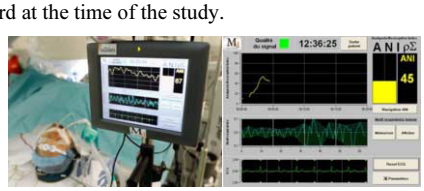


Figure 1 – The PhysioDoloris® monitor in operating room

The HF/E intervention [3] consisted of (i) collecting the mandatory information to prepare the usability plan, (ii) performing the usability verification and validation and (iii) documenting the Usability Engineering File.

HFE experts point of view

An inspection of the Usability Engineering File provided by the HF/E experts to the manufacturer was performed to analyze the type of UEP implemented (steps and methods), the terms used in the document and the reported use error risks. All those elements were compared with the intended classic UEP in the domain of HF/E (ISO 13407) and the intended UEP in the MD regulation (IEC 62366).

Manufacturer point of view

An inspection of the documents provided by the manufacturer to the HFE experts at the beginning of the intervention was performed (accompanying document and first version of the risk management file). The objective was to analyze the usability items already documented by the manufacturer before the HFE experts intervention with regard to the intended IEC 62366 requirements (the application specifications (with the intended user profiles, the intended conditions of use, etc.), the frequently used function, the hazards and hazardous situation related to usability, etc.). Then, the last version of the file given for CE marking by the manufacturer was analyzed to identify the way the items of the Usability Engineering File provided by the HFE experts were integrated: what items have been included? Were they modified and, if yes, how?

Results

Analytical review of the standard document

References

The results show that the ISO 14971 (i.e. the RMP) is the founding reference of the IEC 62366. On the one hand, 93.4% of the references of the normative part refer to the ISO 14971 (the remaining 6.6% concern the IEC 61258 which outlines a generic process for developing materials for education and training for medical electrical equipment). On the other hand, the ISO 14971 is the only referenced document cited as indispensable for the application of the IEC 62366 (Article 2. Normative references).

Table 1 – Distribution of the standards quoted in the IEC 62366 bibliography section by field of expertise

Areas of expertise	Standards cited in the Bibliography section
Quality management system	ISO 9000:2005; ISO 9001:2000; ISO 13485:2003; ISO/TR 16142:2006; EN 1041:1998
Basic safety and essential performance	CEI 61258:1994; ISO/CEI 51:1999; CEI 60601-1:2005; CEI 60601-1-8:2006
Usability (HF/E)	ISO 9241-1:1998; ANSI/AAMI HE48:1993; ANSI/AAMI HE74:2001

The analysis of the Bibliography section shows that references to standards of the HF/E domain are almost non-existent. A total of 75% of the standards cited in the bibliography refer to other fields of expertise as HF/E (Table 1): 5/12 rely on quality management expertise, 4/12 on basic safety and essential performance and, only 3/12 on HF/E expertise.

Terms and defintions

These results are confirmed by the vocabulary listed in Article 3 of the IEC 62366 which relies mainly on other fields of expertise as HF/E. Only 2 terms explicitly refer to standards of the HF/E domain (i.e "effectiveness" to the ISO 9241-11 and "user interface" to the ANSI/AAMI HE74). All other terms (46.2% corresponding to 12/26 terms) refer to standards of other fields of expertise as HF/E (i.e. 9/26 on basic safety and essential performance standards, 2/26 on QMP standards and 1/26 on RMP standards).

Twelve terms have no specific references meaning that they are defined by the IEC 62366 itself. It has to be noted that among these 12 terms, some of them (4/12) have been modified from usual definitions of the HF/E domain: “Usability”, “Use error”, “User” and "Efficiency". For instance, the definition of "Usability" is quite different from the ISO 9241-11 definition (Table 2).

Table 2 – Comparison between the ISO 9241-11 and the IEC 62366 for the definition of the term "Usability"

IEC 62366 (Article 3.17)	**ISO 9241-11** (Article 3.1.)
« *Characteristic of the user interface that establishes effectiveness, efficiency, ease of user learning and user satisfaction* »	« *Extent to which a product can be used by **specified users** to achieve **specified goals** with effectiveness, efficiency and satisfaction in a **specified context of use*** »

Usability Engineering Process

The UEP objectives described in the IEC 62366 differ slightly from those of the ISO 13407. The classic UEP aims to support the design of interactive systems ensuring their usability on the whole while the usability MD regulation aims to identify and prevent risks of use errors. IEC 62366 adopts a safety point of view while the classic usability approach is more global.

This discrepancy obviously impacts the requirements of each standard. The main difference is that the ISO 13407 stresses the importance of understanding and specifying the context of use on the whole (Figure 2), i.e. considering the intended users, their tasks with the device and the corresponding work organization. In contrast, the IEC 62366 emphasizes the need to describe the specifications of the application, the frequently used functions and the identification of hazards and hazardous situations related to usability, getting closer to the risk analysis step of the ISO 14971. Accordingly, the ISO 13407 sets usability objectives related to users and organizational requirements for the evaluation step while the IEC 62366 imposes usability objectives related to frequent and hazardous functions.

Concerning the intended level of requirements, the normative part of the IEC 62366 presents a UEP process similar to the UEP process described in the ISO 14971. The UEP as described in usual HF/E standards (ISO 13407 & ANSI/AAMI HE74) is in an informative section of the IEC 62366 (Annex D).

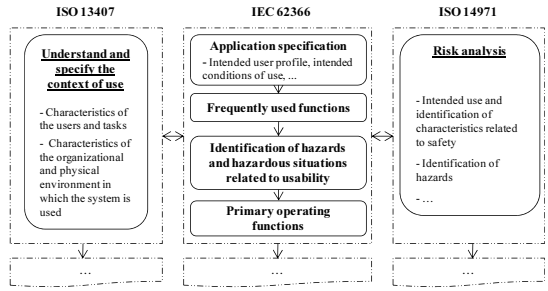


Figure 2 – Description of the first steps of the processes described in the IEC 62366, ISO 13407 and ISO 14971

Analysis of the discussion forum

Twenty-eight threads were included in the analysis corresponding to 295 posts. ALCESTE classified 563 ECU out of 793 created. It revealed 5 stable lexical classes which were pooled into 2 lexical meta-classes (Table 3).

In the first meta-class (45.65% of ECU), the participants want to understand how to reach compliance with the HF/E regulatory requirement. Two lexical classes are included in this meta-class. The first one (Class A, 23.09% of ECU) represents discussions where participants give information about harmonized standards (role and status). It highlights difficulties of the participants to identify the suitable usability standard regarding their MD. In the second class (Class B, 22.56% of ECU), difficulties are expressed with the understanding of the IEC 62366, participants asked for feedbacks and trainings on this purpose.

Table 3 – Lexical meta-classes and classes and examples of typical lexical features

Lexical meta-class (ECU %)	**Lexical classes (ECU %)**	**Examples of typical lexical features**
1. Searching the compliance with the ergonomics regulatory requirement (45.65%)	A. Provide information on compliance with the regulatory requirements in general and especially with the usability harmonized standards (23.09%)	iec6060116, standard<, harmonized, adopted, solution+, why, mandatory, regulation+
	B. Search for information and help to understand the IEC 62366 standard (22.56%)	help, thanks, committee, publication, feedback, follow, training, experience, provide+, us, hope
2. Understanding the application of the IEC 62366 (54.35%)	C. Guidance in understanding the UEP principles (27.35%)	user+, design+, error+, interface+, verification, HF, environment+, pati+ent, test+, population+, clear+, real
	D. Search for information about the documentation of the UEF and its link with the RMP (19.36%)	file+, usability, engineering, template+, document+, manage+, plan, existing, procedure, include, risk+, require+, process+, our+

	E. Search for information about the RMP methods to identify risks of use errors (7.64%)	DFMEA, analysis+, detectability, risk+, hazard+, covered, failure+, chang+er, determine+, incorrect+, control+

The second lexical meta-class (54.35% of ECU) includes discussion aiming to understand the application of the IEC 62366 principles. Three classes are part of it. The Class C (27.35% of ECU) includes explanatory utterances about the objectives, the scope and the definitions of the IEC 62366. The Class D (19.36% of ECU) highlights difficulties of participants with the documentation of the UEF. The items "file+" and "document" are the most typical items of this class and one of the most repeated utterance is "usability engineering file" showing that participants are focused on the UEF documentation. Participants try to understand the benefits of the UEP as they don't grasp the relationship between UEP and RMP processes, e.g. how are these two processes different from each other. For example, a representative ECU of this class is: *"Annex of iec14971 takes usability into account, [...] etc. can iec62366 requirements just be implemented in the overall risk plan for the product, or will the auditor want to see a separate usability specification like annex h?"*

Finally, the last class, Class E (7.64% of ECU), includes utterances discussing the methods of the RMP that can be used to identify and prevent the risks of use errors expected by the IEC 62366. The typical lexical features highlight (i) a risk management vocabulary as "hazard", "failure", control", "severity" and "risk" and (ii) a vocabulary related to methods as "determine", "list", "detectability", "analysis", "organized", "method". The primary methods cited are "FMEA" and "FTA", typical risk management methodologies. For example, the most representative ECU of this class is the following: *"My question is whether or not it is recommended to use detectability in an FMEA to comply with iec62366?"* In this class, the ISO 14971 annexes are regularly cited as good guidance to identify use error.

Application of the standard : case study analysis

HFE experts point of view

It has to be noted that the IEC 62366 does not require a specific format for the Usability Engineering File.

In this case study, the results show that the HF/E experts had adopted a classic UEP with comprehensive analysis of the context of use and the evaluation of the designed solutions. Likewise, the documentation of the Usability Engineering File was done following a usual usability report. Although some terms of the IEC 62366 were used (e.g. application specification, frequently used functions, etc.), HF/E experts mainly used specific terms of the classic UEP (e.g. context of use, heuristic analysis, usability testing, etc.).

In regards to the methods, one main difficulty emerged. The HF/E experts have had some difficulties in defining thresholds for the usability goals as they were not used to doing that exclusively in a safety-oriented point of view. These criteria for determining adequacy to the requirements referred to the criteria defined for risk acceptability in the ISO 14971. Moreover, they didn't distinguish in the usability engineering file the safety-oriented usability aspects of the problems, rather they were related to ease-of-use as required by the IEC 62366.

Manufacturer point of view

Analysis of the documentation recovered from the manufacturer shows that only one intended user profile was documented, i.e. the anesthetist, while another major user of the MD was the nurse. In France, indeed, an anesthetist is often in charge of 5-6 patients inside an operating room while nurses are assigned to a given patient. Nurses are also the ones monitoring the patient state and even making some decisions, although obviously the physician is the final decision maker. The manufacturer knew this specific work organization, but in order to be able to control risks associated with this additional user profile, he had decided at the time to consider only the physician's profile as the physician was legally the only one responsible for the drugs administered to a patient. At the end of the study, the nurse profile was added into the documentation by the manufacturer.

Another major finding was that the manufacturer only focused on technical risks (e.g. related to electrical problems). Even if they knew the problems identified by the HF/E experts, they had not identified them as risks of use errors. For instance, Figure 3 illustrates a severe usability problem identified by the HF/E experts generating a risk of misinterpretation of the index which was potentially dangerous for patient safety.

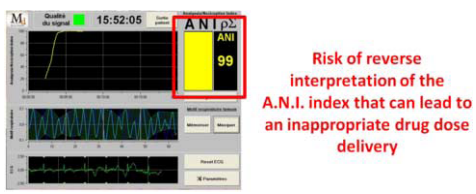


Figure 3 – Example of a dangerous usability problem identified by HFE experts

From the manufacturer's point of view, it was not a risk as the choice of the direction of the index relied on a clinical explanation (the A.N.I. indicates the proportion of parasympathetic tone in the autonomous nervous system). According to the manufacturer, once the care providers have understood it (during MD training), it would no longer be a problem. He did not consider that training was a low-level risk control measure while an easy high-level countermeasure could have been implemented with the redesign of the Graphic User Interface (GUI). They had difficulties in linking a design choice of the GUI with a potential risk of use error.

Discussion/Conclusion

This study highlights the need of multiple expertises to be able to understand and apply the IEC 62366 standard; it especially shows the need of a double expertise: the HF/E and the risk management expertises. The results show that the QMP expertise is not essential to grasp the content of the IEC 62366 as the standard does not really rely on the ISO 13485.

Firstly, even if the IEC 62366 standard claims to be a usability standard, it has little explicit elements about usability (few references to HF/E, key usability elements only in informative parts). Moreover, it clearly looks like the ISO 14971standard (a lot of references to it and a normative part dedicated to the description of a process very close to the RMP).

Secondly, these results are confirmed by those of the discussion forum analysis. Participants have difficulty understanding the distinctive feature of the IEC 62366 since, for them, the process is similar to the process of the ISO 14971. They also tend to interpret the IEC 62366 based on the risk management process which make easier things for them. As the risk management requirements have been mandatory since the 1990s, it is a process already systematized for most of the manufacturers and thus, well-known by them.

Thirdly, the analysis of the case study reveals that the manufacturer does not really understand the usability-related risks as shown by their not considering the nurse's user profile. As in many French operating rooms, nurses have a major role in the process of anesthesia, and as such cannot be ruled out. A problem would be considered as a use error and not as an abnormal use (i.e. outside the manufacturer's obligations).

Finally, problems of misinterpretation of the IEC 62366 were also observed with the HF/E experts. Spontaneously, they have adopted a classic UEP and have not emphasized the safety-oriented point of view of the identified usability problems (no clear distinction of the usability problems linked to risks of use errors).

Based on all these results, it seems that the IEC 62366 is a usability standard presented and structured as a risk management one. It requires the RMP expertise to be able to grasp the issues of the risk analysis and to master the related methods, but it also requires HF/E expertise to be able to correctly apply the UEP with the identification and prevention of the major risks of use errors.

The Health Information Technology (HIT) community must today be attentive to the MD regulation which has a strong impact on the software MD design and development. The major risk with the IEC 62366 standard lies in the problem of misinterpreting the requirements without realizing because each person understands the standard based on his/her own expertise. The recent focus on patient and user safety related to HF/E aspects makes the usability engineering aspects in development, implementation, and use of medical software a key issue that requires the development of good practice guidelines and standards in this area.

Standards are designed by several international or national standardization organizations (e.g. ISO or IEC) involving many technical committees. HF/E standards are developed and published by these standardization groups [10]. In the HF/E domain, the International Ergonomics Association (IEA) initiates in 1974 was the first HF/E technical committee (TC 159) for the ISO. Today, every standardization organization has a technical committee dedicated to HF/E. But it seems that the international Standard IEC 62366 has been prepared by a joint working group of three committees (subcommittee 62A: Common aspects of electrical medical equipment used in medical practice; IEC technical committee 62: Electrical medical equipment in medical practice and technical committee; ISO/TC 210: Quality management and corresponding general aspects for medical devices) not by integrating HF/E committees. Moreover, the technical committees are usually composed of manufacturers and customers [11]. One action to be taken is to ensure that at least domain experts understanding the issues discuss them within the committees.

From the practical and organizational perspective, we need to make the HF/E and risk management expertises and guidance more visible and accessible to the IEC 62366 intended users and to support all forms of education and training of stakeholders for HIT and MD.

Acknowledgments

Part of this research was funded by the ANR-11-TECS-0011 grant. We wish to thank MDoloris Medical Systems® for sharing their time and expertise.

References

[1] Council Directive 2007/47/CE, Council Directive 2007/47/CE, (2007).

[2] International Electrotechnical Commission, IEC 62366:2007 Medical devices – Application of usability engineering to medical devices, Geneva, 2007.

[3] Bras Da Costa S, Pelayo S, Bastien C, Jeanne M. Beuscart-Zephir MC. (2011). The new standard for medical devices' usability and safety: a serious research issue? Communication presented at the 5th International Symposium on Human Factors Engineering in Health Informatics, Trondheim, Norway, August 26-27.

[4] Thovtrup H, Nielsen J. Assessing the usability of a user interface standard, In S. P. Robertson, G. M. Olson, & J.S. Olson (Eds.), Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems, ACM, New Orleans, Louisiana, pp. 335-341, 1991.

[5] International Organization for Standardization, ISO 13485:2003 Medical devices – Quality management systems – Requirements for regulatory purposes, Geneva, 2003.

[6] International Organization for Standardization, ISO 14971:2007 Medical devices – Application of risk management to medical devices, Geneva, 2007.

[7] International Organisation for Standardization, ISO 13407:1999 Human centred design processes for interactive systems, Geneva, 1999.

[8] ANSI/AAMI HE74:2001. Human factors design process for medical devices. Arlington, VA.

[9] Reinert M. Un logiciel d'analyse lexicale. Cahiers de l'analyse des données. 1986;11(4):471–81.

[10] Vanderdonckt J. Development milestones towards a tool for working with guidelines. Interacting with Computers. 1999; 12(2): 81–118.

[11] Dul J, De Vries HJ. Differences between ergonomics and other ISO/CEN standards. Lessons for standards development. In *Proceedings of the 16th world congress on ergonomics*. Maastricht, The Netherlands, 2006.

Address for correspondence

Centre d'Investigation Clinique et d'Innovation Technologique de Lille, Maison Régionale de la Recherche Clinique, 6 rue du professeur Laguesse, 59045 Lille Cedex, sabrina.brasdacosta@univ-lille2.fr.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-358

Evidence-based Heuristics for Evaluating Demands on eHealth Literacy and Usability in a Mobile Consumer Health Application

Helen Monkman[a], Janessa Griffith[a], Andre W. Kushniruk[a]

[a] *School of Health Information Science, University of Victoria, Victoria, BC, Canada*

Abstract

Heuristic evaluations have proven to be valuable for identifying usability issues in systems. Commonly used sets of heuritics exist; however, they may not always be the most suitable, given the specific goal of the analysis. One such example is seeking to evaluate the demands on eHealth literacy and usability of consumer health information systems. In this study, eight essential heuristics and three optional heuristics subsumed from the evidence on eHealth/health literacy and usability were tested for their utility in assessing a mobile blood pressure tracking application (app). This evaluation revealed a variety of ways the design of the app could both benefit and impede users with limited eHealth literacy. This study demonstrated the utility of a low-cost, single evaluation approach for identifying both eHealth literacy and usability issues based on existing evidence in the literature.

Keywords:

Consumer health informatics; usability; eHealth literacy; health literacy; heuristic evaluation; mHealth

Introduction

Potential patient safety implications arise from the use of consumer health applications (apps), and consumer health information systems (HISs) more generally. As such, new strategies should be explored to identify potential problems and resolve them. To this end, the National Health Service (NHS) currently offers a "Health Apps Library" which endorses apps that are considered to a) be relevant to British people, b) provide trustworthy information, c) abide by data storage regulations, and d) not pose potential risks due to improper use [1]. The clinical assurance team, comprised of doctors, nurses and safety specialists, review apps and collaborate with app developers to ensure clinical safety standards are met before they are accepted into the Health Apps library [1]. However, the review appears to focus solely on guided user decision-making without the oversight of a healthcare professional, as the only potential clinical risk consumer HISs can generate. Although this is a promising and important new strategy, other factors may create clinical risk through the use of consumer HISs. For example, incongruence between the demand the system places on eHealth literacy and the user's actual eHealth literacy skills could result in misinterpretation of information.

Health Literacy is the "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [2]. However, technology has the potential to introduce entirely new challenges for consumer seeking health information health literacy stemming from system interactions. Emphasizing this argument, eHealth Literacy is "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [3]. Moreover, it is not prudent to investigate the demand placed on eHealth literacy by a system without considering the system's usability more generally, as either factor being suboptimal may deter or prevent consumer HISs use.

Heuristic evaluation is a popular usability inspection method because it is a rapid, low-cost, investigation that often provides useful insights to improve system usability. Heuristic evaluation could be considered as a complementary and/or preliminary method of identifying potential usability issues that can be remedied before investing in more expensive and time consuming usability testing with representative users.

Although commonly used sets of heuristics (e.g., Nielsen's 10 [4]) are typically applied, these generic sets of heuristics may not always be the most suitable for a particular evaluation, user group, or system. Thus, some researchers have opted to develop their own heuristics based on the specific goal(s) of an assessment. For example, heuristics have been developed to evaluate health information system safety [5].

Another shortcoming of commonly applied sets of usability heuristics is that they predominantly focus on the analysis of the software component and neglect the system content or information presentation. It has been previously argued that special design considerations are necessary for users with limited health literacy, as they often use systems differently than more health literate users [6]. This argument led to the development of new heuristics, potentially more suitable for consumer health information systems [6]. However, a limitation of the previous work [6] is that the heuristics were generated from a single resource: *Health Literacy Online: A guide to writing and designing easy to use Web sites* [7]. Albeit composite and thorough, Health Literacy Online was published in 2010; since its publication, there has been an influx of research on eHealth/health literacy and usability. Thus, there is an opportunity to refine heuristics initially proposed by Monkman and Kushniruk [6], as well as generate new heuristics based on the recent literature.

This study will test the utility of a new set of evidence-based heuristics derived from the literature on eHealth literacy and usability. The heuristics will be used to evaluate a mobile blood pressure app to determine how using this app might benefit consumers, as well as what obstacles this app might present to users in terms of both demands on eHealth literacy and usability.

Methods

Heuristics and Severity Scale

Building on previous work [6], the current investigation sought to expand on the heuristics orginally proposed by the

authors, by incorporating recommendations from additional research on eHealth/health literacy and usability. Therefore, the body of literature investigating eHealth/health literacy and usability in conjunction were used to generate a novel set of heuristics. Barriers and recommendations reported in the literature were then subsumed into heuristics. This analysis identified a total of 8 essential heuristics (see Table 1) with an additional 3 heuristics relevant to specific types of content or medium that can be used dependent upon the system under investigation (see Table 2). This set of heuristics were then applied to test their utility assessing a mobile consumer health application to identify potential issues related to demands on eHealth literacy and usability.

Table 1– Evidence-Based Heuristics for Health Literacy and Usability

	Heuristic	Description
1.	**Immediately Inform Users of Purpose and Engage Users, Avoid Registration**	Identify the purpose and audience on the home screen/page. If unavoidable, make registration and logging in simple and obvious.
2.	**Use Complementary Interaction Methods**	Make use of alternative inputs (e.g., touch screen, barcode scanning, voice commands) and outputs (e.g., audio recordings, videos, text to speech engines).
3.	**Leverage Interactivity**	Offer interactive tools (e.g., quizzes, questionnaires, glossaries, tutorials) to engage with the information and provide performance feedback. Allow users to share information (e.g., print, email) with others.
4.	**Provide Accurate, Colloquial, Comprehensive, Succinct Content**	Written information should be brief, relevant, and in users' vernacular.
5.	**Provide Tailored, Flexible, Layered Content**	Prioritize information according to importance. If possible, personalise informtion. Provide succinct summaries but allow users to access more detailed information. Offer content in multiple languages.
6.	**Use Visuals to Complement Text, But Avoid Tables**	Visuals (e.g., pictures, videos, animations) may enhance written information. If unavoidable, tables should should be designed as independent, simplistic representations of information
7.	**Simplistic, Consistent Navigation**	Keep users oriented. Use linear navigation to facilitate forward and backward movement. Use large buttons, clearly labeled links, and provide a search engine.
8.	**Simplistic, Consistent Displays**	Avoid on screen complexity. Avoid the need for scrolling by limiting information on a page / screen.

Table 2– Optional Evidence-Based Heuristics for Health Literacy and Usability For Specific Content or Device

	Heuristic	Description
9.	**Clear and Comprehensive Communication of Risks**	Describe risk terminology in ways users will understand. Use 100 as upper limit on bar graphs. Avoid logarithmic scales.
10.	**Clear Depiction of Monitoring Data and/or Test Results**	Emphasize values outside acceptable ranges. Facilitate pattern recognition and rapid identification of influential factors.
11.	**Considerations for Mobile Devices**	Allow users to adjust the display size using familiar input (e.g., pinch to zoom, turning to landscape orientation). Use appropriately sized interface elements. Limit the amount of information displayed.

The severity scale used for this evaluation was specifically developed for rating health literacy or usability issues identified in consumer HISs [6]. The three severity levels (i.e., mild moderate, severe) are used to [6]:

1. Prioritize issue resolution
2. Estimate the likelihood consumers will understand the content and the gravity of the consequences associated with misunderstanding
3. Indicate the extent to which users will be able to circumvent the obstacle posed by the issue

System Under Evaluation

A mobile blood pressure tracking application (app) was selected as the system for this evaluation. This app operates on iOS systems (i.e., iPhone, iPad, iPod) and the full version is available for $0.99 CAD. Although its primary purpose is to record blood pressure values over time, this app also tracks heart rate and weight. This app was selected because it is an affordable solution that may appeal to people with limited ehealth literacy who are at risk or have been diagnosed with high blood pressure.

Procedure

In addition to the eight essential heuristics, the optional heuristics, Clear Depiction of Monitoring Data and/or Test Results and Considerations for Mobile devices, were both relevant to this system and therefore applied in this evaluation.

Given that the mobile blood pressure tracking app aimed to facilitate the monitoring of data over a time period, the evaluation focused on two stages of app usage: 1) the initial profile generation, and 2) reviewing data trends. As such, the investigators populated the app with several blood pressure measurements, both in and out of the acceptable range. This approach is recommended for tracking/monitoring apps, as it makes the evaluation more representative of what users would view after using the system multiple times.

Two usability experts with no clinical expertise (HM and JG) used the heuristics while performing tasks to investigate a blood pressure app independently with the goal of identifying strengths and weaknesses of the application. For investigations aiming to identify potential issues with eHealth literacy, it is advisable that the evaluators do not have clinical expertise, as they may be better able to detect potential content issues (e.g., terminology, undefined acronyms) for representative users.

Results

Favourable Aspects of the Blood Pressure Tracking App

The blood pressure tracking app provided opportunities to share the data users entered (e.g., email, print).

The app color coded text and data point values that were considered out of range leveraging the convention of green for normal values, yellow/orange to indicate prehypertension, red for hypertension, and blue to denote hypotension.

This app offered several different views of the entered data: a) a summary table depicting maximum, minimum, and average values; b) a graph of the data; c) a chart of the descriptive statistics; and d) a frequency table of how many values fell within each range. The date range (i.e., day, week, month, year) for each report is adjustable and also allows showing all values, or only those taken in am or pm. Individual entries can be modified in the history.

The majority of the information in this app was kept above the fold, eliminating the need for scrolling and minimizing the likelihood of missed information.

This app also offers the additional security feature of setting a passcode lock and allows users to backup and restore data through both WiFi and iCloud.

Opportunities to Improve the Blood Pressure Tracking App

For brevity, this section will only outline violations that were deemed moderate or severe, organized according to heuristic. The heuristic Clear and Comprehensive Communication of Risks was not evaluated and therefore not included here.

A total of 40 heuristic violations were identified in this analysis. This blood pressure tracking app had the most issues associated with its complex navigation (9) and display (8) (see Figure 1). Unfortunately, the majority of the violations identified were considered either moderate (18) or severe (17) in nature, either because they were insurmountable usability problems or they had the potential to misinform users. In the interest of brevity, only the violations deemed to be severe will be discussed. Both investigators identified 14 violations. The remaining unique issues were found by one investigator (HM = 23) or the other (JG = 3).

Immediately Inform Users of Purpose and Engage Users, Avoid Registration

The app forces the user to make a profile to enter data. It is not clear what values are mandatory. The app applies default values for birthday, weight, height, gender, goal blood pressure, goal heart rate, goal weight, which is not prudent. There is an explanation of the purpose of the field labeled “color range”, which would likely confuse users. If multiple profiles exist in the app, the active profile is not obvious, which could result in adding data to the wrong profile.

Use Complementary Interaction Methods

This app fails to use any alternative methods to input or output information, which increases the probablilty of inaccurate data entry due to manual error. Additionally, the majority of values are entered using “pickers” (scrollable menus), which maybe be less efficient and more tedious than providing a numerical keyboard for data entry.

Leverage Interactivity

Although this app enables users to share their tracking data, it offers CSV, HTML, and PDF formats but does not describe why a user would select one file type over another. Additionally, this app does not engage users by offering any interactive learning tools or resources on blood pressure.

Provide Accurate, Colloquial, Comprehensive, Succinct Content

This app offers very limited content and therefore does not facilitate understanding the underlying mechanisms of blood pressure nor influencing factors. Unfortunately, the app uses multiple undefined acronyms (e.g., BP, HR, MAP, mmHg, bpm), which may be confusing for some users.

Provide Tailored, Flexible, Layered Content

Despite forcing users to create a profile, default goal values do not vary as a function of what is entered as current values nor clinical guidelines (e.g., using BMI for goal weight). As well, alerts are not provided when the user enters a value that would be considered outside of a healthy range. For example, entering a systolic value of >180 mmHg, which is considered to be a hypertensive crisis [8], does not produce an alert. Moreover, the app fails to incorporate additional information that is clinically relevant for blood pressure monitoring (e.g., whether a user is pregnant or has diabetes).

Use Visuals to Complement Text, But Avoid Tables

Very few visuals were used to complement text information. The few icons displayed were used suboptimally and felt busy, because they were often intermixed with data. Unexpectedly, the camera button captures a screenshot rather than a photo.

Unfortunately, multiple tables were used to display data, which would likely present challenges for users attempting to extract information from them.

Simplistic, Consistent Navigation

The app allows users to navigate through different date ranges; however, when a user navigates to date ranges too far in the past or future, there is no simple way to get back to the most current entries. The button to add a new entry is displayed in the top right corner on most pages, but is replaced with a camera button on some pages. The buttons to add weight, notes, or change the date of an entry are not obvious, which could cause users to become disoriented. Additionally, users are able to enter data for dates in the future, which is inadvisable. Further, it was not obvious that there were multiple different reports, as one of the investigators nearly missed the additional summaries entirely.

Adding a new notification, or reminder to take blood pressure was unnecessarily complex and had a different navigation pattern than other features. Though this is not a critical function of the app, it could be beneficial in facilitating consistent tracking by users.

Simplistic, Consistent Displays

The colors in the app are generally distracting (e.g., gradients, shadows) and low contrast. There are multiple lines (both solid and dashed) on the graphs which provide limited value and increase the impresson of a cluttered display.

When entering a value, no labels denote the systolic, diastolic, and heart rate on either the input component or display of entered information. Additionally, new entry dates are displayed only as numbers, a potentially confusing format.

There were three significant display inconsistencies of note. First, the scales on the graph change when different time spans are selected, presumably according to the range of values on the display. Second, the format of the PDF summary of values that can be shared by email were very different from any of the displays within the app. Finally, red, orange, green, and yellow are used inconsistently. For data points, they indicate thresholds within acceptable ranges, yet on other displays these colors are used strictly for differentiation and do not reflect whether the value entered is in or out of range. This inconsistency could potentially result in misinterpretation.

Clear Depiction of Monitoring Data and/or Test Results

The app uses orange for the systolic line on the graph and green for diastolic. This could create confusion if users interpret that the green diastolic line is always within range.

Although the information is presented in several ways, the tables and charts may be difficult for users to interpret and extract information from, and could potentially confuse them.

Considerations for Mobile Devices

Unfortunately, this app does not allow users to change the size of the font either by reorienting the mobile device or in the settings. Additionally, some of the buttons were small.

Other

Errors are likely to occur because the app automatically enters information. Specifically, the app saves its default values for blood pressure, heart rate, and/or weight values for any new entries, if they are saved without modification.

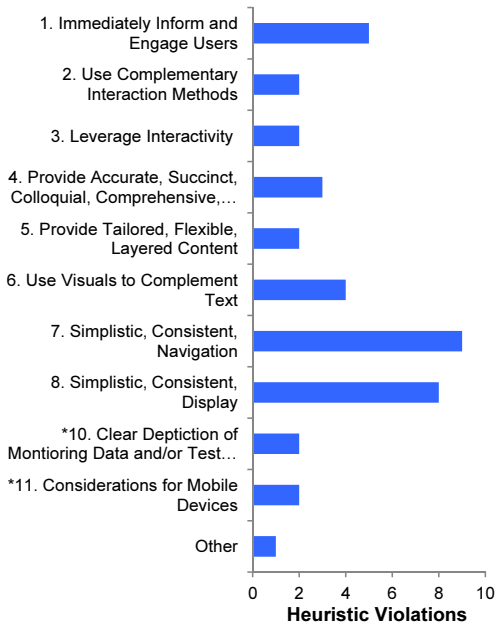


Figure 1 – Frequency of Violations as a Function of Heuristic
** Denotes optional heuristics relevant for this evaluation.*

Proposed Revisions for the Blood Pressure Tracking App

Figure 2 was developed to depict design changes to improve the app, particularly for users with limited eHealth literacy, by resolving some of the issues identified in the heuristic evaluation. Generally, a more minimalistic, higher contrast appearance was adopted, to both emphasize the information and enhance readability.

Given that adding values is a primary task, this function was promoted to having its own button on the menu. Thus, users could easily add a new entry from anywhere within the app. Although this app did not connect with blood pressure cuffs, it is always preferable that where possible, data is automatically uploaded into consumer health applications but allows users to revise or complement data with additional information (e.g., a note). Additionally, as shown in Figure 2a, the proposed redesign provides a holistic view of all of the information that can be added to a new entry. In contrast, the existing app forces users to toggle between screens to add different types of information. This display also shows the labels and units of the information, unlike the existing app. Finally, icons are used to reinforce the data labels.

Several recommendations were used to simplify the graph display (see Figure 2b). Specifically, the acceptable range of values for systolic and diastolic blood pressure were superimposed onto the graph in light green to emphasize values in and out of range. This color coding scheme was continued further, by redundantly coloring acceptable values in green, while hypertensive values were orange or red depending on severity, and hypotensive values were colored blue. Additionally, it is advised that the color coding scheme adopted would be applied to summary tables and individual entries. To prevent disorientation, the revised display constrains users to viewing the graph with today as the most recent entry to prevent disorientation.

It is also recommended that the app would incorporate a glossary of terms and/or links to trusted references and educational resources in the tools (e.g., [8]) to facilitate users understand the concept of blood pressure, related terminology, and what factors affect blood pressure.

(a) (b)

Figure 2 – Proposed Revised Displays for the Blood Pressure Tracking App

Discussion

Currently, the blood pressure tracking app has limited value to consumers due to the numerous moderate and severe potential usability and eHealth literacy demand problems identified here. It is not to say that users were not be able to use this system for monitoring, but it would not provide them with an enhanced understanding of what blood pressure is and what behaviours influence it. Users could also potentially misinterpret information or rely on default values. However, design modifications and the incorporation of actionable content could increase the usefulness and usability of this app.

This novel set of evidence-based heuristics derived from studies designing or evaluating consumer HISs for users with limited eHealth literacy, demonstrated utility in revealing potential problems Specifically, this evaluation revealed a number of opportunities to improve the blood pressure tracking app to improve its usability for consumers with limited eHealth literacy and how to more effectively communicate information to these users. The issues identified here stemmed from design issues with either the interface or the content that could be problematic for users with limited eHealth literacy. Moreover, given that users of these systems often have minimal or no clinical expertise, it is imperative that information is written and displayed in a manner that scaffolds understanding.

Leveraging findings from existing literature on usability and eHealth literacy to develop heuristics for inspecting other consumer HISs is a promising technique. Specifically, this method of developing evidence-based heuristics facilitates

mitigating known barriers to both use and comprehension for users with limited eHealth literacy with a single evaluation. Applying these heuristics could help ameliorate system content and design prior to full-scale usability testing.

This study expanded on previous work by incorporating additional evidence on usability and eHealth literacy. Our initial heuristics [6] were more specific, whereas the heuristics presented here were broader. There are advantages and disadvantages to using either set. Investigators with less familiarity with demands on eHealth literacy and usability issues may find it more helpful to use heuristics that provide specific guidance. Whereas, evaluators with more experience in these domains may find it preferable to use more general heuristics, as they are more flexible. This rationale may account for the differences in problem identification between inspectors in this study, as one inspector has more experience in this domain. Moreover, the heuristics presented are still preliminary, as they are still not entirely mutually exclusive and require more refinement. Further, as the body of research on eHealth literacy and usability continues to mount, emerging evidence should be continuously integrated into this set of heuristics.

The findings from this study suggest that additional heuristics are needed to classify all identified problems. That is, the "Other" category of problems had to be created to accommodate problems that were not subsumed in one of the existing heuristics. This suggests that the developed heuristics might benefit from being complemented with other heuristics such as Nielsen's [4] ten heuristics. Amalgamating the set of heuristics generated here with Nielsen's [4] could provide a more comprehensive set of principles for designing and evaluating consumer HISs is thus one proposed line of future research.

There are several other opportunities for further research using these heuristics. First, these heuristics may be useful for comparing and contrasting multiple consumer health information systems with similar purposes (e.g., health risk assessments, blood glucose monitoring apps) in an attempt to determine the most suitable of these systems for consumers with limited eHealth literacy. Another line of research that warrants investigation is testing the validity of these heuristics by comparing the problems found through their application with problems revealed through usability testing of the same system, to determine the correlation between the their respective findings.

Conclusion

As the proliferation of consumer HISs continues, it is important to consider the usability and the demands these systems place on eHealth literacy. Some might argue that low cost applications should not be scrutinized, given their restricted resources. However, users with limited eHealth literacy are equally, if not more, likely to use these inexpensive systems. Thus, it is important to recognize and apply low-cost techniques, such as the one outlined in this paper, to improve consumer HISs of all varieties. Moreover, these heuristics facilitated the identification of both eHealth literacy and usability problems with only a single evaluation.

Consumer HISs have the potential to engage and empower new user groups to monitor and improve their health. That is, these systems could possibly increase users' participation in their own healthcare by providing personalized feedback and tailored interventions that cater to the needs and preferences of their users. However, for this information to be received and applied by the users who have challenges using technology and difficulty understanding health information, it is imperative that they are designed with usefulness, usability, and eHealth literacy considerations.

Although heuristic evaluation with evidence-based heuristics may be valuable for informing the design of consumer HISs to mitigate potential problems for users with limited eHealth literacy, it is not meant to replace usability testing with representative users. Instead, such heuristic evaluation should be used as a preliminary filter to identify and resolve issues commonly experienced by limited literacy users. Thus, it would facilitate development of more usable applications, as less time would be spent repeatedly identifying problems through usability testing that are known problems, previously cited in the literature. That is, using evidence-based heuristics (i.e., design principles for users with limited eHealth literacy) as a preliminary method of improving a consumer HIS, should result in more effective use of representative users finding problems unique to a specific consumer HISs under investigation. Thus, this method should be used in conjunction with other user-centred design methods as a part of an exhaustive, iterative design process to ensure that consumer HISs are easy to use and users understand the information contained therein.

References

[1] NHS. Choices. Health Apps Library. About. Retrieved December 1, 2014 from http://apps.nhs.uk/about/

[2] Ratzan, S. C. & Parker, R. M. (2000). Introduction. In National Library of Medicine Current Bibliographies in Medicine: Health Literacy. NLM Pub. No. CBM 2000-1. C. R. Selden, M. Zorn, S. C. Ratzan, and R. M. Parker, Editors. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services. Retrieved November 30, 2014 from http://www.nlm.nih.gov/archive//20061214/pubs/cbm/hliteracy.pdf

[3] Norman, C. D., & Skinner, H. A. (2006). eHealth literacy: Essential skills for consumer health in a networked world. Journal of Medical Internet Research, 8(2), 0-0. doi: 10.2196/jmir.8.2.e9

[4] Nielsen, J. Heuristic evaluation. In: Nielsen J and Mack RL, eds. Usability Inspection Methods. John Wiley & Sons: New York, 1994. pp. 25-62.

[5] Carvalho, C. J., Borycki, E. M., & Kushniruk, A. W. (2009). Using heuristic evaluations to assess the safety of health information systems. Studies in Health Technology and Informatics, 143, 297-301.

[6] Monkman, H., & Kushniruk, A. (2013). A health literacy and usability heuristic evaluation of a mobile consumer health application. Studies in Health Technology and Informatics, 192, 724.

[7] U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Health literacy online: A guide to writing and designing easy-to-use health Web sites. Washington, DC. 2010.

[8] American Heart Association. Understanding Blood Pressure Readings. Retrieved December 4, 2014 from http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp

Address for correspondence

Helen Monkman, School of Health Information Science, University of Victoria, Victoria, BC, Canada, email: monkman@uvic.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-363

Evaluating the Impact of Player Experience in the Design of a Serious Game for Upper Extremity Stroke Rehabilitation

Marcos Cordeiro d'Ornellas[a], Diego João Cargnin[a], Ana Lúcia Cervi Prado[b]

[a] *Laboratório de Computação Aplicada, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*
[b] *Hospital Universitário de Santa Maria, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*

Abstract

Video games have become a major entertainment industry and one of the most popular leisure forms, ranging from laboratory experiments to a mainstream cultural medium. Indeed, current games are multimodal and multidimensional products, relying on sophisticated features including not only a narrative-driven story but also impressive graphics and detailed settings. All of these elements helped to create a seamless and appealing product that have resulted in a growing number of players and in the number of game genres. Although video games have been used in education, simulation, and training, another application that exploits serious gaming is the exploration of player experience in the context of game research. Recent advances in the natural user interfaces and player experience have brought new perspectives on the in-game assessment of serious games. This paper evaluates the impact of player experience in the design of a serious game for upper extremity stroke rehabilitation. The game combines biofeedback and mirror neurons both in single and multiplayer mode. Results have shown that the game is a feasible solution to integrate serious games into the physical therapy routine.

Keywords:

Computer Games; Patient Engagement; Rehabilitation; Stroke.

Introduction

Regardless of the advanced medical technology and highly trained healthcare professionals, individuals under medical supervision do not always follow their treatment courses resulting in decreased adherence. Solutions to this problem are complex with behavioral and psychological factors playing an important role. A recent review of video games' application to improve healthcare outcomes reported therapeutic benefits, especially in the psychological and physical therapies [1].

Video games seem simple entertainment applications, but their use could also open entirely new frontiers. The term serious game has arisen to describe the use of games in education, training, health, and public policy [2]. Serious games modify a learning task into the one that is game-like, which deeply reworks the player experience. Describing such player experiences has become an important branch of research.

Games are used in healthcare to engage individuals and to improve their health outcomes. Several studies have shown positive results with the use of serious healthcare games. This could be because the game environment is aesthetically appealing, graphically engaging, and provides the necessary motivation to endure treatment with joy and less pain.

The technologies for natural user interfaces (NUI), and the rehabilitation tools development environment for the physically disabled individuals, are well documented in the literature [3]. Serious health care games can replace many of the rehabilitation exercises that impaired individuals could otherwise perform in virtual environments. Besides, the physical therapist can shape the treatment by configuring various features of rehabilitation exercises while evaluating and controlling the performance.

For the individuals with motor disabilities due to stroke or aging, performing the rehabilitation exercises is essential while they still have some motor skills [4]. Individuals may start the rehabilitation process by getting involved in the serious healthcare games. The serious healthcare gaming experience is about producing the required stimuli while keeping it amusing. Therefore, the player experience and engagement are the mechanisms that make serious games motivating and fun.

This paper presents a systematic evaluation of player experience in a serious healthcare game designed for upper limb stroke rehabilitation. The game is tested in single- and multiplayer modes on a control group of physical therapists to measure intragroup interactions following a quantitative analysis. A one-way ANOVA is used to compare game modes with respect to challenge, competence, flow, negative effect, positive effect, immersion, tension, functional limitations, and perceived exertion.

This paper is structured as follows: Related Work, The Game, Experimental Evaluation, Results, Discussion, and Conclusion and Further Work. In the 'Related Work' section, the paper presents information concerning serious games and motion capture devices, the importance of biofeedback and mirror neurons in rehabilitation, and the comments on player experience in serious games. 'The Game' section introduces the game platform involving upper extremities' rehabilitation after stroke. The 'Experimental Evaluation' section provides initial setup information, measuring instruments, participants, and procedures. The 'Results' section presets the reasonable outcomes that are hypothesized before conducting a statistical analysis. The Discussion section analyzes the results, and the 'Conclusion and Further Work' section presents meaningful observations while outlining the work for future experimental activities.

Related Work

Serious Games and Motion Capture Devices

Graphical user interfaces are ubiquitous on personal computers and mobile devices. Designers face new challenges from the current user interface technology and applications that integrate touchscreens, 3D depth cameras, and gesture and voice recognition. These interfaces allow user communications with a variety of devices that imitate the real word interaction.

The development of a widespread range of exetreme games has resulted in the use of NUI devices for physical therapy. For example, it is now feasible to measure therapy progress for individuals with motor disabilities, who are instructed to perform exercises, irrespective of their location (home or clinic). A compiled database of regular exercises with quality software must be created to assure flexibility and to support the development of tailored rehabilitation therapies for individuals suffering from a particular disease.

The game industry provides tracking sensors with a broad range of applications and reliability for rehabilitation purposes [5]. Low cost, off-the-shelf availability and response characteristics of these devices have enabled their adoption for the healthcare applications. Their relevance to the individuals with motor disabilities attending therapy sessions is discussed in [6]. Though the extreme games help improve healthcare outcome they were not known to physical therapists until recently and therefore were not used for rehabilitation purposes.

The popularity of Kinect – a motion capture device created by Microsoft – promoted the emergence of several studies exploring its application to serious games for healthcare [7], including those for physical and rehabilitation therapies. A wide range of serious healthcare games has been developed for physical therapy [8]. Serious game-based therapy aims at encouraging individuals with motor disorders to practice physical exercises, however, some of them lack access to the cloud computing applications [9] while others do not have significant technical information that is valuable during physical therapy assessment [10].

Biofeedback and Mirror Neurons in Rehabilitation

The biofeedback is adopted into a bio-inspired user interface. The healthcare applications use integrated biofeedback sensors to process vital bio-signals [11]. Biofeedback measures user experience and engagement in serious healthcare games [12]. Data acquisition is one of the first steps in the evaluation and applicability of serious games throughout the treatment.

Biofeedback involves conscious and visual control of a particular activity. Biofeedback-assisted training aims at developing full control over physiological processes, it helps shaping the posture and profile during rehabilitation and recovery. Biofeedback, among other things, consists of altering sound intensity and visual colors in accordance with the task. Individuals learn to control the activity as per the information acquired from the feedback loop.

Functional damage due to cerebral palsy requires rehabilitation and recovery. Rehabilitation must begin soon after the stroke. Physical therapists may provide detailed diagnostics before the therapeutic approach. Care must be taken regarding neurological and neuropsychological deficits that affect the treatment and length of stay at a rehabilitation facility. For this evaluation, post-stroke monitoring includes clinical tests and modern methods of measurement, such as mirror therapy.

Player Experience in Serious Games

Literature on evaluation of serious games focuses on learning outcomes. Although quantitative estimates are prone to heuristics and biases, it is important to understand them and to work out ways to mitigate their influence on the evaluation. One possible solution is to measure user experience with qualitative or quantitative surveys, and questionnaires. They are amongst the most widely used methods in gathering information on subjects, providing feedback to the player based on actual gameplay data [13].

Experience occurs in response to a stimulus. The stimuli (in general) result from observation or participation in real, imaginary, or virtual events. Experiences are not spontaneous phenomena but are induced, and have a reference (starting point) and intention (aim for something) [14]. Although the concepts of reward and comfort channels are game-related, they must be accounted when considering a player's experience.

The Game

The serious game was developed with scenes that change with the game mode. The control panel lets a physical therapist set the goals in accordance with the required treatment. The game features single and multiplayer modes, as shown in Figures 1 and 2, respectively. In single player mode the patient plays alone reaping as many fruits from the tree as possible over a certain period of time. In multiplayer mode, one may choose to have a competitive or collaborative game. Participants reap fruit towards achieving a target (of a selected number of fruits), which might be specified by the physical therapist. Participants are evaluated by a digital goniometer, displayed on both sides of the screen. A multimedia database is included to store game sessions for game experience assessment, including mirror-neurons and biofeedback.

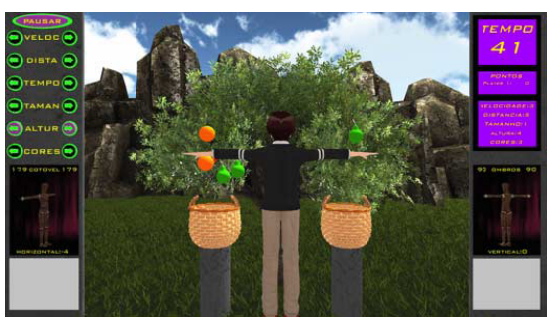


Figure 1 – Screenshot of the serious game for upper-limb stroke rehabilitation in single player mode.

Figure 2 – Screenshot of the serious game for upper-limb stroke rehabilitation in multiplayer mode.

Experimental Evaluation

Initial Setup

The step-by-step instructions for initial setup for the serious game users are listed below:

1. Participants get a brief explanation about the game metaphor and the data gathering process. The trainer instructs participants about the information and consent form needed for clinical trials.
2. Participants must be in the sensor range to play the game. Participant positioning depends on whether the game is played in single- or multi-player mode.
3. Exercises and procedures are taught during a first trial. Evaluation starts only after the participants have their questions clarified and they are acquainted with the game.
4. Game session setup is common to all of the participants. Game sessions are video-recorded for later analysis.
5. The game session lasts until the preset time limit. Also, the participants are instructed to fill out online questionnaires.

Measuring Instruments

The Game Experience Questionnaire (GEQ) is designed to measure different aspects of players' engagement and skills in playing a serious healthcare game. Research has shown the importance of players' social connections to the overall player experience [15]. The GEQ allows game researchers to investigate and test player experience of serious games in a quantitative and inclusive manner [16].

The GEQ analyzes the user experience across various phases of the game and has three modules: the core module, the social presence module, and the post-game module. The core module addresses competence, sensory and imaginative immersion, flow, tension, challenge, and favorable and adverse effect; the social presence module deals with empathy, negative feelings, and behavioral involvement; and the post-game module focuses on positive and negative experience, tiredness, and returning to the reality. Only the core module was used in this research to assess the game experience that included forty-two items. Scores are computed based on the Likert scale.

Two additional measuring instruments were used to assess and record player experience during the gaming sessions. 'Capabilities of the upper extremity' questionnaire (CUE-Q) [17] was applied to measure improvements after upper extremity reconstructive procedures. CUE-Q measures functional limitation, and assesses the amount of difficulty experienced in performing specific actions with one or both arms and hands. Although a measure of functional limitation in tetraplegia, it may also be used to measure how well people respond to a treatment. The Borg rating of perceived exertion (RPE) [18] is a widely used measurement for monitoring and guiding exercise intensity. Despite being a subjective measure, RPE provides valuable information and is sufficiently accurate in capturing the experience pertaining to effort and workload.

Participants

Thirty-eight adults participated in the evaluation. None of the participants were diagnosed with motor disabilities. Participants came from a specific control group constituted by physical therapy students, therapy technicians, physical therapists, and professors. Participants were recruited on a voluntary (no monetary compensation) basis from the rehabilitation ward of the HUSM.

Procedure

The experiment took place at the neuro-rehabilitation ward of the HUSM. The room where experiments were performed has plenty of space for in-game assessment. It also provided controlled illumination, and thermal and acoustic insulation against interferences that might distract the participants. Being at the rehabilitation ward was not a matter of choice, but a necessity. It enabled shorter evaluation sessions and a quicker feedback during the development.

Positioning of devices within the room was driven by the sensor coverage range. The screen was installed on the wall, raised to a comfortable position and providing enough side-to-side space for participants. The Kinect was installed just below the screen. Trainers and physical therapists, while leading the experiment, did not interfere with the participants' performance, but could instruct participants, observe participants' progress, and control the game.

Evaluation data is stored in a database for each participant, including left and right shoulder abduction angles, left and right elbow flexion angles, and horizontal and vertical slopes.

Results

This research aims to evaluate player experience induced by a single player mode against a multiplayer collaborative game mode within the control group of physical therapists. Research hypotheses for the experiment on the data gathered from serious game sessions are :

- H1: Reported levels of engagement will be higher for players in multiplayer mode than in single player mode.
  - Player experience will vary with challenge, competence, and flow.
  - Multiplayer mode will be more intense with respect to sensory and imaginative immersion, resulting in higher engagement scores.
  - Positive effect and tension will be higher in multiplayer mode than those in single player mode. Likewise, negative effect will be lower in multiplayer mode than that in single player mode.
- H2: Measured levels of functional limitation and the amount of difficulty experienced in performing specific game actions will be comparable in both game modes. The same holds for perceived exertion.
- H3: Multiplayer condition provides not only a great deal of fun but also provides an associative learning experience through biofeedback and mirror neurons.
- H4: Participants in either (single- and multi-player) modes will perceive a positive impact on how they interpret their progress. Thus, local and online game scores and achievements will induce an enhanced performance.

A one-way ANOVA was used to determine whether player experience differed based on the GEQ core module, CUE-Q, and RPE, amongst the users of single- and multi-player modes. A one-way ANOVA is used to compare the means for *n* independent groups on a dependent variable. The F-statistic and its p-value depend on the group means and on the spread of the observations within each group. A few values in the

dataset were missing or outliers. Therefore, before the analysis, average values of measures were normalized using a logarithmic transformation – a standard procedure.

As to the challenge in the single player mode, a higher level of difficulty and effort required by the serious game was obtained (items 8, 26, 29, and 37), which indicated that the game was challenging and stimulating while being a learning activity. Time pressure was observed during sessions. A majority of participants agreed that the game was not difficult nor was significant effort needed (items 13 and 36). An equivalent interpretation was observed for multiplayer mode.

The sensory and imaginative immersion that investigates a player's presence in the game indicated a higher percentage of agreement regarding all but item 20, which pertains to the imaginative feeling. These results indicated that a majority of the participants found the game interesting and the serious game provided a creative and rich experience. Single- and multi-player modes did not indicate significant differences concerning immersion. The tension scores, which measured the level of frustration or annoyance, were highly positive implying that the control group of players did not feel tense, annoyed, irritable, or frustrated while playing the game. Average tension was somewhat higher in single-player mode than that in multi-player mode.

Scores on GEQ were measured only for the core module on a 1 (not at all) to 5 (extremely) scale. Both game modes showed high scores ($M > 4.0$) on the items pertaining to positive effect and immersion. Median values were observed for challenge, competence, and flow (M in [2.5,3.5]), while negative effect and tension were rated low ($M < 1.8$) as seen in Figure 3.

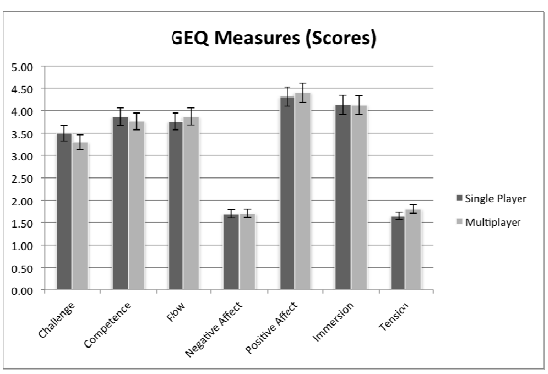


Figure 3 – Comparison of score for GEQ dimensions in single player and multiplayer mode.

Results of the ANOVA reveal that differences between the single- and multi-player modes are more significant regarding challenge. Challenge is high in single player mode with $F(1,74)=6.1715$, $p=.005$. Since $F>F^{Crit}$, the mean challenge of at least one game mode is significantly different. The difference in challenge is statistically significant between single player and multiplayer mode. Mean squares were also higher in single player than in multiplayer mode. Large F values were observed within challenge measurements. A large F-ratio occurs when sampling had large values in some groups and small values in others. All the other GEQ dimensions followed the null hypothesis-driven approach (the population means for both game modes are the same) when evaluating player experience on single player vs. multiplayer mode.

Running a multivariate ANOVA on all seven dimensions showed no significant difference on positive effect, and sensory and imaginative immersion perceived in either mode. This suggests that single player participants enjoyed the game as much as their multiplayer partners. Significantly low scores were found for negative effect and tension. No significant differences were found between single player and multiplayer mode for CUE-Q and RPE.

Discussion

Metaphoric introduction of characters in the game for either game modes may have given participants something familiar to connect with, enhancing feedback, and allowing them to feel more immersed in the serious health game experience. By focusing on the goals, participants had to remain more concentrated on the game, adding to the flow component.

Participants felt more distracted in the single player mode while they had thought about other things in the multiplayer mode. Other than preventing wrong movements by the real-time sensor, audio was not used to its full potential. Sounds and/or soundtracks may be relevant when evaluating negative effects.

Challenge measurement provided a wider range of results when compared to other GEQ dimensions. Although the game metaphor, characters, and interface were the same for either game modes, the very same action or decision choice of collecting fruits from the trees had an unexpected outcome and the challenge is directed by the effort needed to complete the task that does not differ in both game modes.

A majority of the hypotheses were confirmed by the results, while there were some unpredicted results. Challenge, competence, and flow varied in both game modes. A small but significant difference was observed regarding player experience in both player modes with respect to the sensory and imaginative immersion. As expected, positive effect and tension scored higher in multiplayer mode. Conversely, negative affect was almost equal in both game modes. Although average scores on negative affect were low, neither game mode seemed to annoy or frustrate participants.

Measured levels of functional limitations, and perceived exertion, were similar in both game modes. The high scores confirmed hypothesized effects of biofeedback and mirror-neurons on game experience for single player and multiplayer modes. Apart from tension, no significant differences were found. The user-centered design process in both game modes may support these findings.

Conclusion and Further Work

We showed an evaluation of the player experience with a serious game for upper extremity stroke rehabilitation. To assess player experience during gameplay it is essential to consider both the design and simulated environment for the participants while assessing the immersive interactions they undergo in single player and multiplayer modes.

Evaluation participants fit the target user group of physical therapists. Disagreements and misconceptions were observed between participants that created divergences. Open conflicts during gameplay provide means to solve a particular problem while misconceptions might generate negative feelings between participants. Intragroup variability was detected and evaluated using analysis of variance (ANOVA) techniques. Additionally, the measuring instrument must also be appropriate to qualitatively assess and identify experiment outcomes. Engagement was evaluated by means of GEQ dimensions, CUE-Q, and RPE.

Results indicated that engagement were somewhat higher in multiplayer mode (collaborative) than in single player mode

due to the in-game control, which facilitated learning by means of biofeedback and mirror neurons. The same results produced high levels of challenge, competence, flow, and positive affect but also a sense of tension and distraction. The analysis revealed that participants in multiplayer mode had a better understanding of the collaborative tasks.

Apart from limitations due to the amount of participants in the control group and game attributes with desired learning effects for serious games (which can be accessed directly during gameplay) differences between single player and multiplayer mode indicated the presence and influence of behavioral states on player experience. Game developers and physical therapists agreed that, despite the user-centered game design, it was possible to highlight the differences more precisely and in a timely manner.

Further work is needed to assess engagement while monitoring emotional responses in real-time for the treatment group and the control group. Such a research will help the serious healthcare games' designers over the course of development and will provide useful biofeedback to improve player experience.

References

[1] Primack BA, Carroll MV, McNamara M, Klem ML, King B et al. Role of Video Games in Improving Health-Related Outcomes: A Systematic Review. American Journal of Preventive Medicine 2012;42(6):630-638.

[2] Wouters, P., E. van der Spek, and H. Van Oostendorp, Current practices in serious game research: A review from a learning outcomes perspective. Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices, 2009: p. 232-250.

[3] Meleiro P, Rodrigues R, Jacob J, and Marques T. Natural User Interfaces in the Motor Development of Disabled Children, Procedia Technology, Volume 13, 2014, Pages 66-75

[4] Chang YJ, Chen SF, and Huang JD. A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. Research in developmental disabilities, 32(6):2566–2570, 2011.

[5] Tanaka K, Parker J, Baradoy G, Sheehan D, Holash J et al. A comparison of exergaming interfaces for use in rehabilitation programs and research. Loading... North America, 6, feb. 2012.

[6] Deutsch JE, Borbely M, Filler J, Huhn K, and Guarrera-Bolby P. Use of a low-cost, commercially available gaming console (wii) for rehabilitation of an adolescent with cerebral palsy. Physical Therapy, vol. 88, no. 10, p. 1196–1207, 2008.

[7] Krammer J, Burrus N, Echtler F, Herrera DC, and Parker, M. (2012) Object Modeling and Detection. In: Hacking the Kinect, Apress, Berkeley.

[8] Pastor I, Hayes H, and Bamberg S. A feasibility study of an upper limb rehabilitation system using kinect and computer games. In: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012, pp. 1286–1289.

[9] Fraiwan M, Khasawneh N, Malkawi A, Al-Jarrah M, Alsa'di R et al. Therapy central: On the development of computer games for physiotherapy. In: Innovations in Information Technology (IIT), 2013 9th International Conference on, 2013, pp. 24–29.

[10] Sugarman H, Weisel-Eichler A, Burstin A, and Brown R. Use of novel virtual reality system for the assessment and treatment of unilateral spatial neglect: A feasibility study. In: Virtual Rehabilitation (ICVR), 2011 International Conference on, June 2011, pp. 1–2.

[11] Jercic P, Astor PJ, Adam MTP, and Hilborn O. A Serious Game Using Physiological Interfaces For Emotion Regulation Training In The Context Of Financial Decision-Making. (2012). ECIS 2012 Proceedings. Paper 207.

[12] Frutos-Pascual M, Zapirain BG, and Buldian KC. Adaptive cognitive rehabilitation interventions based on serious games for children with ADHD using biofeedback techniques: assessment and evaluation. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '14), Brussels, Belgium, 321-324.

[13] Obrist M, Roto V, and Väänänen-Vainio-Mattila K. User experience evaluation: Do you know which method to use? In: Proceedings of CHI 09. Boston, MA, 4-9 April 2009. New York, NY: ACM Press. 2763-2766

[14] Nacke, LE, Drachen A, Korhonen H, Kuikkaniemi K, Niesenhaus J et al. Playability and Player Experience Research. September 1, 2009. Proceedings of DiGRA 2009: Breaking New Ground: Innovation in Games, Play, Practice and Theory, Brunel University, West London, UK.

[15] Nacke L, and Lindley CA. Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience. In: Proceedings of the 2008 Conference on Future Play: Research, Play, Share, Toronto, Canada, ACM, pp. 81-88.

[16] Gajadhar BJ, Kort YA, and Ijsselsteijn WA. Shared Fun Is Doubled Fun: Player Enjoyment as a Function of Social Setting, Proceedings of the 2nd International Conference on Fun and Games, October 20-21, 2008, Eindhoven, The Netherlands.

[17]Oleson CV and Marino RJ. Responsiveness and concurrent validity of the revised capabilities of upper extremity-questionnaire (CUE-Q) in patients with acute tetraplegia. Spinal Cord. 2014 Aug;52(8):625-8.

[18] Borg G, Hassmen P, and Lagerstrom M. Perceived exertion related to heart rate and blood lactate during arm and leg exercise. Eur J Appl Physiol Occup Physiol 1987;56(6):679-85.

Address for correspondence

Marcos Cordeiro d'Ornellas
ornellas@inf.ufsm.br
Av. Roraima, 1000
Prédio 7 – Anexo B- Sala 388
UFSM - Campus Universitário
97105-900 Santa Maria-RS – Brasil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-368

User-Centered Design of Health Care Software Development: Towards a Cultural Change

Enrique Stanziola[a], María Quispe Uznayo[a], Juan Marcos Ortiz[a], Mariana Simón[a], Carlos Otero[a], Fernando Campos[a], Daniel Luna[a]

[a] Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina

Abstract

Health care software gets better user efficiency, efficacy and satisfaction when the software is designed with their users' needs taken into account. However, it is not trivial to change the practice of software development to adopt user-centered design. In order to produce this change in the Health Informatics Department of the Hospital Italiano de Buenos Aires, a plan was devised and implemented. The article presents the steps of the plan, shows how the steps were carried on, and reflects on the lessons learned through the process.

Keywords:

User-Centered Design; User-Computer Interface; Software Design.

Introduction

Health professionals and patients value health care software, but they demand better products that help them satisfy their clinical information needs [1]. According to reviews, many of the issues that prevent adoption and satisfactory use are related to poor design and low usability [2, 3, 4]. Usability is a measure of the efficacy, efficiency, and user satisfaction [5], and usability has been established as a key factor of applications that provide the needed support while letting them be focused on their tasks [6]. Usable applications are easy to learn, efficient to use, easy to remember, not prone to errors, and subjectively pleasing to use [7]. User-centered design (UCD) is an approach to achieve usable products introduced by Norman and Draper [8]. UCD has evolved from the research field of human-computer interaction, with the contibution of cognitive psychology, software engineering, sociology, and other disciplines that influenced it and developed methodologies that embody their points of view [9] Usability engineering presents an iterative cycle of design and evaluation until the fulfilment of established quantitative goals measured on lab tests [10]. Contextual design includes ethnographic methods for user research while participatory design involves users to set goals and explore design solutions, instead of just taking part of evaluation [11]. UCD has been applied succesfully to enhance adoption and success of products [12].

The Health Informatics Department of the Hospital Italiano de Buenos Aires (HIBA) includes an area of software engineering and a residence for health informatics. The department has been designing and developing administrative and health care software since the late 1990s. However, these applications presented the usability challenges described in the literature mentioned above. While the users valued the availability of online health records and tools for administrative tasks, they also noted lack of desired functionality and difficulties to learn and use the tools. Therefore, acoording to the reccomendations on the cited reviews, the department decided to implement UCD, applying usability engineering, participatory design, contextual design and a user-centered framework for redesigning health care software [13].

The goal of this paper is to describe the process of changing the culture of design and development of health care software at the HIBA, setting UCD at its core.

Methods

Following Schaffer's approach to institutionalization of usability [14], the head of the department devised a plan of three phases.

The first phase was to introduce UCD through an internal course for health informatics residents and software developers and other personnel of the department. The course load was 16 hours of classes, plus extra activities.

The second phase was to demonstrate the feasibility of UCD, applying it to a pilot project. The team for the project would be taken from the attendees of the course that show more interest and developed skills.

The third and current phase is aimed to the widespread implementation of UCD on the mainstream projects of clinical software. It was planned to create a UX team, formed by a group of UX professionals. It has a double function: to offer the service of designing new applications or new versions of existing ones, and to disseminate UCD as a new culture of work.

Results

UCD introductive course

In January 2011, the planned course of UCD was given and it was mandatory for health informatics residents, developers and other personnel from the department. It was a program of 1 class of 2 hours per week, for 8 weeks. User interviews, prototyping and usability testing were required as practical activities outside the class.

Although approximately 30 people attended this course, its impact was mixed since very few developers showed interest on usability. The resistance arguments were that: (1) users have to adapt and learn; (2) health domain is complex and is impossible to make it easier; and (3) UCD would be the right thing to do, but is not possible given the conditions of work in the department.

On the other hand, most of the health informatics residents welcomed the new approach centered on users. But, they usually stay 2 years on the department and new residents have to participate on UCD courses every year. Another challenge

is their few design skills, so they have to learn almost from scratch.

Pilot project

After the course, a team for a project was taken from the attendees of the course that show more interest and developed skills. The team was formed by a health informatics resident, a MD student, and a part-time user experience (UX) professional. Its goal was to redesign the HIBA's patient health portal and it was a test of UCD techniques. That project involved real patients for interviewing and testing usability and proved the worth of this approach by giving new insights that helped to focus the development. The team took a transdisciplinary approach, looking forward that all of its members could develop UCD skills and view health issues from an assistive technology perspective. The result was a revised version of the application, that included a new menu structure, functionality and visual style. Findings of this project are shown on Table 1. After that pilot project, UX professionals were hired by the hospital in a full-time basis and are taking part in several projects.

The UX team

Since January 2014, the User Experience (UX) team at the HIBA is formed by 3 professionals with different skills that complement each other. Its leader was formed and worked in information systems analysis and development, and later studied human-computer interaction. Two members were chosen and they worked on graphic design, usability and accessibility. Their job is to lead the UCD process and promote that methodology among their colleagues in the department. In this way, UX members plan and perform tasks of user research, prototype design and usability evaluation, along with HI residents and developers. Non-UX professionals take part on basic training given by the UX team, and this training allows them to understand the UCD's point of view and its techniques. Due to high demand of projects, each member works on several projects at a time, and delivers different levels of attention for each one. Before implementation, applications get at least design advices and revisions by the UX team. Priority projects get the most attention, which involves thorough user research and testing.

Impact of UCD on real projects

After three years of the formation of the UX team, we can see its achievements on the design of health care software. Table 1 lists the projects that got some contribution from the team, the insights found through user research and its impact on the final design.

Mobile usability lab

The core of the UX team's work is to listen and observe what users say and do on interviews, workshops and tests. So, it is necessary to record their words and actions to be analysed afterwards. The team uses small cameras and cellphones to take pictures, videos and audio on contextual interviews. The team uses a notebook for on-location usability tests. The first notebook used for tests was a Dell Latitude E5510, with a 15.6' screen and 1,366x768 pixels resolution. Its display was smaller than screens installed in doctor's and nurse's offices, so it was replaced by a another notebook, a Dell Precision

Table 1- Projects with contribution from the UX team

Project	Findings of user research	Impacts on the interaction design
Patient health portal [15]	The application is valued even by users with IT challenges. Registering and navigating was difficult. Patients with chronical diseases, including senior citizens, are the most intensive users.	The menu was revised using card sorting technique with participants profiled on age and chronical diseases. A style guide was created, taking into account recommendations for older users.
Problem selection [16]	On problem search, relevant results act as suggestions for selecting more specific problems.	The list of results was ordered by relevance and classified by categories to stimulate the selection of more precise problems.
Anesthetic record [17]	Understanding how anestethiologists record their actions during an surgical intervention.	Interaction design of touch-screen based application created from scratch. Interactive and real-time visualization of vital signs and administerd medication.
Evolutions and documents [18]	While writing evolutions, doctors usually read results and other data. Doctors usually look for evolutions written by colleagues of their specialty. Documents were difficult to find.	A floating window that allows to navigate the EHR while writing evolutions was designed. Faceted browsing was added, including filters by type of document and specialty.
Drug-drug interactions alerts [19]	It is difficult to understand and follow recommended actions embedded on text.	Specific buttons to perform recommended actions were added, customized for each drug-drug interaction case.
Nursing process [20]	Large and complex forms hinders completion of tasks.	Forms were rearranged according to the hierarchy of items to minimize perceived complexity and completion times.

M6800 with a 17.3' and 1,920x1,080 pixels, to avoid distortion of prototype and similar legibility as the user's monitors.

The notebook is equipped with Morae 3.2 software to record the user's face expressions, the application on display, the mouse movements, and the keyboard typing.

Discussion

The plan to adopt UCD in the department has achieved its first goals after 4 years. The projects implemented show quality improvements that were welcomed by sponsors and users. This methodology is now valued and demanded by health informatics residents and software developers. Below, we discusss specific considerations about the lessons we learned in the process.

Teaching UCD in organizations to change the software development culture

We found out some reasons behind the resistance faced during the course. As developers had almost no free time for classes, and many of them were not initially motivated to learn UCD, we believe that elective and shorter courses could have better results.

Health is effectively a complex domain and it is difficult to reach a usable and useful design. UCD instructors have to be careful and avoid putting themselves in the place of 'gurus' that know exactly how to do the right thing, while developers do not. Instead, before teaching UCD it would help to expose developers with real users interacting with the applications and expressing their difficulties. Thus, developers would learn from their users that UCD is necessary for developing applications that deal better with the genuine complexity of health domain. This is based in an essential principle of UCD: to know the users [21]. In our case, there is a great variety of users, regarding their profiles, demographics and roles: patients, clinicians, surgeons, nurses and other health professionals. For each project listed on Table 1, the user's capabilities, needs and goals were inquired. The profiling of these users can be consulted on the respective cited publications.

It is necessary to know more about how projects are carried on at the organization before proposing a new process of design. Learning UCD is, above all, learning a new way to work. For that reason, the instructor, along with the course attendees, have to find out the necessary steps to get from the current situation to the desired one. The know-how has to be constructed, not imposed, if we want it to be embraced rather than resisted.

UCD as service

UCD became accepted when developers recognized its valuable contribution to their projects. The UX team helps the work of developers by designing the interaction and the user interface. Now, they can focus more on the difficulties of systems, database and communications. Users and sponsors show more satisfaction; therefore, developers get more recogniticon for their work. When this virtous circle was evident, developers begin to ask the UX team for help on their projects, even in informal ways.

The key factor for the change in the developers' attitude towards UCD, from resitance to demand, was that the UX team is at their service to work with the developers.

UCD as culture

Progressively, UCD is becoming pervasive, and it is not just the UX team's job. Developers and HI residents are conscious that their goal is to get satisfactory user experiences. The department is getting a common view of UCD as a key of success of its work. This goal was achieved primarily by practicing methodologies and showing good results in user performance and satisfaction. Since the first attempts in teaching UCD faced resistances, the team's main effort was not on telling but showing its value. The change of culture was not immediate, but durable and contagious. Now, very few developers still resist UCD.

Perspectives of UCD for health care software

The practice of UCD for health systems has to be transdiciplinary in order to be viable and effective. Projects are better run by teams formed by staff with diverse knowledge and skills, including medicine, nursing, UX and development. The methodology includes aspects of participatory design, since each member learns the basics of the others' disciplines.

The design of applications should integrate in an overall service that offer a continuous experience for professionals and patients. If not, there is a risk of getting well designed but disconnected touchpoints, that may pose risks as well as dissatisfaction.

Conclusion

This article shows the process taken in the Health Informatics Department of the Hospital Italiano de Buenos Aires to set UCD at the core of the design and development culture, in order to enhance usability of health care software. We carried out a plan that included teaching a course, experiencing a pilot project and forming a UX team for service and dissemination. The article reports how it was implemented and the obtained achievements: impacts on the design of health care software applications, and a positive cultural change in the department towards the adoption of UCD as essential for its practice.

References

[1] Friedberg MW, Chen MW, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. RAND Health, AMA, 2013; pp. 33–46. Available from: http://www.rand.org/pubs/research_reports/RR439.html

[2] Zahabi M, Kaber D, and Swangnetr M. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. Hum factors 2015; Mar 23. DOI 10.1177/0018720815576827.

[3] LeRouge C, and Wickramasinghe N. A Review of User-Centered Design for Diabetes-Related Consumer Health Informatics Technologies. J Diabetes Sci Technol 2013: 77(44): 1039-56.

[4] Hermawati S, and Lawson G. Managing obesity through mobile phone applications: a state of the art review from a user- centred design perspective. Pers Ubiquit Comput 2014: 18(8): 2003-23.

[5] ISO/IEC 9241-11. Geneva, Switzerland: International Organization for Standardization, 1998.

[6] Schumacher RM, Patterson ES, North R, and Quinn MT. NISTIR 7804 Technical Evaluation, Testing and

Validation of the Usability of Electronic Health Records. NISTIR Technical. 2011. Available from: www.usability.gov

[7] Nielsen J. Usability engineering. Boston: Academic, 1993.

[8] Norman DA, Draper SW, eds. User Centered System Design. Hillsdale, NJ: Erlbaum, 1986.

[9] Carroll JM. Human-computer interaction: Psychology as a science of design. Annu Rev Psychol 1997: 48: 61-83.

[10] Nielsen J. op. cit.

[11] Carroll JM. op. cit.

[12] Preece J, Rogers Y, Sharp H, Benyon D, Holland S, and Carey T. Human Computer Interaction. Reading, MA: Addison-Wesley, 1994.

[13] Johnson C, Johnson T, and Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005: 38(1): 75-87.

[14] Schaffer E. Institutionalization of Usability: a step-by-step guide. Reading, MA: Addison Wesley, 2004.

[15] Goldenberg J, Martínez M, Stanziola E, Gómez A, Luna D, and González Bernaldo de Quirós F. Rediseño centrado en el usuario de un portal personal de salud. CBIS. Curitiba, Brasil, 2012.

[16] Luna D, González Bernaldo de Quirós F, Stanziola E, Otero C, López G, and Leguizamón G. Diseño centrado en el usuario para mejorar la carga de problemas en una historia clinica electronica. CBIS. Curitiba, Brasil, 2012.

[17] Sabalza M, Waldhorn M, Bianco G, Garbino J, Stanziola E, Ortiz JM, Garcia G, Luna DR, and González Bernaldo de Quirós F. Desarrollo e implementación de un Registro Anestésico Electrónico propio. Infolac 2014. Montevideo, Uruguay.

[18] García, M, Stanziola E, Ortiz JM, Aguirre Ojea MF, Quispe Uznayo M, López GE, Storani FD, Luna DR, and González Bernaldo de Quirós F. Rediseño del módulo de navegación de documentos clínicos de la historia clínica. Infolac 2014. Montevideo, Uruguay.

[19] Luna DR, Stanziola E, Otero C, Quispe Uznayo M, Belloso W, Risk M, and Gonzalez Bernaldo de Quiros F. Diseño participativo de alertas para la Interacción Droga-Droga en una Historia Clínica Electrónica. Infolac 2014. Montevideo, Uruguay.

[20] Schachner B, Stanziola E, Gassino FD, López GE, Recondo FJ, González ZA, Simón M, Sommer JA, Luna DR, and González Bernaldo de Quirós F. Rediseño de la interfaz de usuario del módulo de enfermería en la historia clínica. Infolac 2014. Montevideo, Uruguay.

[21] Nielsen J. op. cit.

Address for correspondence

Departamento de Informática en Salud, Hospital Italiano de Buenos Aires. Perón 4190 - (C1181ACH) Ciudad Autónoma de Buenos Aires, Argentina.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-372

Online Continuing Medical Education for the Latin American Nephrology Community

Alvaro Margolis[a,b], Francisco Gonzalez-Martinez[c], Oscar Noboa[c], Mario Abbud-Filho[d], Leticia Lorier[a], Marcelo Nin[c], Ricardo Silvariño[c], Sofía García[a], Jacqueline Pefaur[e], Gustavo C. Greloni[f], Irene L. Noronha[g], Antonio Lopez[a,b], María A. Ribeiro-Alves[h], Roberto Tanús[i], Juan Fernández-Cean[j]

[a] *EviMed Corp, Uruguay.* [b] *School of Engineering, Universidad de la República, Uruguay.* [c] *Nephrology Department, Universidad de la República, Uruguay.* [d] *Faculdade de Medicina de São José do Rio Preto, Brazil.* [e] *Facultad de Medicina, Universidad de Chile.* [f] *Nephrology Department, Hospital Italiano de Buenos Aires, Argentina.* [g] *Nephrology Department, Universidade de Sao Paulo, Brazil.* [h] *Nephrology Department, Universidade Estadual de Campinas, Brazil.* [i] *President, STALYC.* [j] *President, SLANH.*

Abstract

A continuing medical education (CME) course was implemented for Latin American nephrologists in 2013. The topic was Immunopathology in native and transplanted kidneys. The course was given in Spanish and Portuguese. The activities included a distance education seven-week asynchronous online modality with multiple educational strategies. Thirty hours of study workload were estimated to complete the course.

Four hundred and ninety-eight physicians coming from 18 countries registered for the course; 442 of them participated in it. Of those who participated, 51% received a certificate of completion and 29% a certificate of participation. Sixty-five percent of registrants participated in the case discussions. Eighty-six percent were very satisfied and 13% were satisfied. Lack of time to devote to the course was the main limitation expressed (62%), while Internet access or difficulties in the use of technology were considered by only 12 and 6% of participants, respectively. There was a significant increase in knowledge between before and after the course; the average grade increased from 64 to 83%.

In conclusion, technology-enabled education demonstrated potential to become an instrument for Latin American nephrologists.

Keywords:

Internet; Continuing Medical Education; MOOC; Nephrology; Latin America.

Introduction

Latin America is a large and diverse region, where the predominant languages are Spanish and Portuguese, comprising more than 600 million inhabitants in over 20 countries, from Mexico in the North to Argentina and Chile in the South.

The number of nephrologists in the region is approximately 8,000, with a widespread variation across countries (1.77 to 53.9 per million population (13.2±14.0)) [1], and also within each of the countries.

SLANH (Latin American Society of Nephrology and Hypertension) is a scientific regional society with over 40 years of existence, integrates 24 Latin American National nephrology associations (www.slanh.net). STALYC (Latin American and Caribbean Transplantation Society) is the transplantation regional federation, with over 30 years of existence (www.stalyc.net). Both joined efforts in this educational initiative.

The use of information and communication technologies (ICT) is evolving constantly, with innovations and the potential for more interactive and engaging strategies to be incorporated into educational programming [2-5], still allowing for massive participation of physicians, some of them from more remote locations of the countries [6]. Therefore, in 2013, EviMed was invited by SLANH and STALYC to implement the integration of ICT into their more traditional formats, based on its record of blended multifaceted programs in Latin America [7-11].

This paper describes the design, implementation and results of a multi-country bi-lingual CME program for Latin American nephrologists.

Methods

The course was implemented in the months of September and October 2013. The topic, Immunology in native and transplanted kidneys, focused on diagnosis, categorization and therapy for these conditions, currently undergoing major discoveries. The target population was Spanish- and Portuguese-speaking nephrologists and other physicians caring for these patients working across the Latin American region. EviMed provided a multidisciplinary team of communication and educational experts, clinicians, system engineers, medical information specialists and translators to design and implement the course, together with regional and international domain experts provided by SLANH and STALYC. There were 33 lecturers, including 25 from Latin America (20 clinical nephrologists and five immunologists), as well as six European and two North American experts. Including the tutoring roles, 59 domain experts participated in the design and implementation of this course.

A multifaceted approach was used, in order to maximize opportunities for physician and health care professional participation and ensure that participants could interact and reflect on the material. Accordingly, the activities began with an on-site and online synchronous launching event, followed

by a distance education seven-week online modality with multiple educational strategies, and ended with a closing lecture. Reading resources, videos and voice-over-presentations were published, together with pre- and post-tests, patient aids, and electronic rounds (e-rounds) on clinical cases. Additionally, a clinical simulation using a custom tool developed with the School of Engineering, Universidad de la República, Uruguay [12], was used to provide applied learning through knowledge discovery and automatic feedback. It was implemented in module 6, almost the end of the course, and it had multiple possible pathways, some better than others, not only right or wrong paths, as in real settings (Figure 1).

The e-rounds were asynchronous discussion groups in either Spanish or Portuguese, coordinated by experts and tutors. This modality allowed physicians to participate whenever they could and wished to, lacking a fixed schedule to access the e-round.

A total of 30 hours of study workload along two months was estimated, in order to complete the course.

Course evaluation, participation, satisfaction and knowledge gain were quantified through pre- and post-tests.

The requirements for the certificate of completion were:

- Access to at least 70% of the study materials.
- Grading of at least 70% of tests in each module.
- Completion of the clinical simulation.
- Participation in at least 50% of discussion forums.

The requirements for the certificate of participation were:

- Access to the virtual campus.
- Access to at least 10 study materials, and/or active participation in clinical discussions.

Wilcoxon signed-rank test was used to compare pre- and post-test results.

Results

Registration

There were 498 physicians coming from 18 Latin American countries who registered for the course (Table 1), along with 59 experts and tutors and 55 observers (total=612 participants).

Table 1 - Distribution of participants by country of origin.

Country	Number	Percentage
Mexico	62	12
Brazil	59	12
Uruguay	53	11
Argentina	48	10
Peru	40	8
Ecuador	32	6
Chile	31	6
Bolivia	27	5
Costa Rica	24	5
Cuba	24	5
Guatemala	23	5
Venezuela	23	5
Colombia	13	3
Dominican Republic	12	2
Paraguay	11	2
El Salvador	8	2
Honduras	4	2
Panama	2	0
Europe	2	0
Total	*498*	*100%*

Regarding the characteristics of the audience, 49% were women and 51% men. Eighty-five percent were nephrologists and 11% were nephrology residents, while the rest were internists, pediatricians, or pathologists, among other. Seventy-seven percent were 34 years or older.

Figure 1 - Example of a graph with multiple intermediate states and paths, in a clinical simulation developed with the mentioned tool.

Participation

Ninety-six percent of registrants (n=479) accessed the online campus, and 92% of them (n=442) participated in the course. Of those who participated, 51% received a certificate of completion (n=226) and 29% a certificate of participation (n=128).

There was a slight decline in participation along the course, as is usually seen in CME courses that occur over a period of time (Figure 2).

The devices used by participants to watch the video-lectures are shown in Table 2. Thirteen percent of accesses were from tablets or smartphones.

Sixty-five percent of registrants participated in the case discussions with peers, tutors and experts, with an average of five participations per registrant out of six forums. Regarding the clinical simulation, 58% of registrants went through it, with an average of 15 pathways explored, out of 33 pathways available.

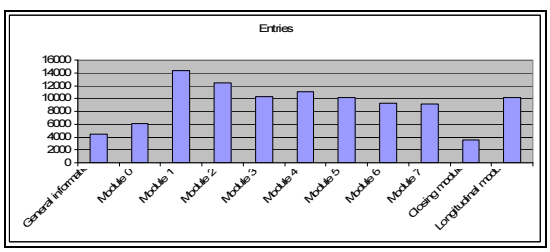


Figure 2 - Number of entries by module.

Satisfaction

This survey was responded to by 174 participants. In a Likert scale of 1-5, 86% were very satisfied, 13% were satisfied, 1% were neutral, and none were dissatisfied or very dissatisfied. Regarding the components of the course that were judged most useful, the video lectures, reading materials, discussion forums and case simulations (in this order) were selected.

Lack of time to devote to the course was the main limitation expressed (62% of participants who answered this question), while Internet access or difficulties in the use of technology were considered by 12 and 6% of participants, respectively.

Regarding perception of commercial bias, less than 1% (1/157) considered there was a commercial bias.

Table 2 - Type of devices used to watch video-lectures

Type of device	Reproductions of video lectures *Number of accesses (%)*	Mean duration (in minutes)
Computer	16,081 (86.2%)	12:03
Tablet	1,951 (10.5%)	11:09
Mobile phone	529 (2.8%)	07:52
Smart TV	57 (0.3%)	07:35
Unknown	46 (0.2%)	05:17

Knowledge acquisition

There was a significant increase in knowledge between before and after the course. Of 200 participants who took both tests, the grade increased from 64 to 83% ($p<0.001$) (Figure 3).

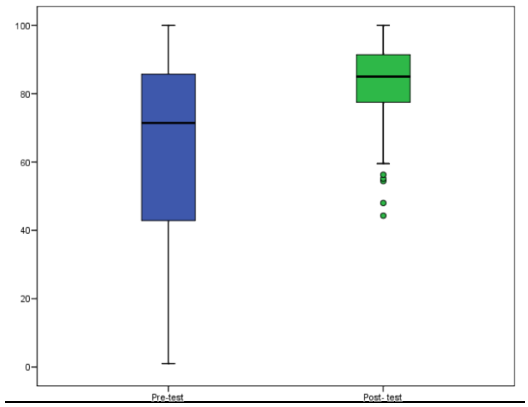


Figure 3 - Knowledge gain.

Discussion

The results of participation and retention shown are aligned with what could be expected in a voluntary online sequential CME program. In particular, the fact that 58% of participants went through the clinical simulation in module 6 is a good indicator of retention. There also was a high level of satisfaction, shown by participants relating to the study materials and experts involved in producing them, the existence of interaction among participants, and the use of innovative educational methods, such as the clinical simulations. The learning gain obtained was significant, as usually occurs with CME based on several features used in this course [13]. The fact that the program was bilingual, and both audiences were integrated in the online platform, received no mentions or criticisms; this is the best possible scenario, where translation was invisible to the user.

This educational activity reached about 7% of nephrologists in Latin America, including faculty and students. In a region as large as Latin America, this kind of program can facilitate participation, no matter the location of the professional.

This initial pioneer program has prompted both SLANH and STALYC to continue using this educational strategy in 2014, with new topics (peritoneal dialysis and solid organ transplantation, respectively). In the case of peritoneal dialysis, the online program for physicians and nurses working in dialysis centers was further integrated with an effort to provide practical support to participants who wish to implement such a dialysis method after the course, from a network of nephrology excellence centers within the region. This kind of strategy increases the chances of implementation of new knowledge and skills learned into clinical practice [14,15].

The cost structure of online education is different from live education [16]. Latin America, a geographically extensive region, but with good internet connectivity and only two main languages, allows for cost savings by avoiding costly transportation and hotel expenses necessary for live events, while engaging participants and experts in a more horizontal and prolonged exchange of experiences.

The two international associations improved their professional networks after the course; in the case of SLANH, its direct membership increased by 50%. This seems to be the result of new communication spaces being created through an online course [17].

The main limitations are related to the fact of having a one-group pre-post test study design, with no control group. Nevertheless, it would not be feasible for these associations to withdraw access to an educational program to a control group. The results regarding enrollment, participation, satisfaction and knowledge gain are useful in themselves, for this given context. Moreover, long-term knowledge retention was not measured, this would be a useful variable to consider in future research. Additionally, there was no formal attempt to measure satisfaction in those who did not answer the survey, who may differ in their perspective related to the course. Finally, no clinical outcomes were measured after this educational intervention; however, this is the norm and not the exception in CME, due to the difficulties related to measuring physician performance or patient outcomes in relation to a course.

A challenge for future courses in Nephrology and for other specialties is the possibility of massive participation of professionals, in the number of thousands, not hundreds, since Latin America has approximately one million physicians. This course had tutors who moderated group discussions, and if the same model is applied with very large audiences, the number of tutors may become difficult to manage. The models used by Massive Open Online Courses (MOOCs) [6] could then be applied here, where either more automatic interaction (e.g., with more use of clinical simulations), and/or more horizontal exchange (e.g., with more peer-to-peer discussions) would lessen the number of faculty required to moderate the discussions.

Conclusion

Technology-enabled education demonstrated good potential to become an instrument for nephrologists as well as other groups of physicians and healthcare professionals working in Latin America. It could help reduce heterogeneity in access to human resources training across regions in the countries.

Acknowledgements

The authors wish to thank the Boards of SLANH and STALYC for their support of this project, as well as the faculty involved in the implementation of the program. We also wish to thank Dr. Alejandro Ferreiro for the statistical analysis.

References

[1] González-Martínez F; Cortés Sanabria L; Di Bernardo JJ; Di Rienzo P; García García G; Cerdas Calderón M; Correa Rotter R; Fernández Cean J. Población y distribución de nefrólogos en Latinoamérica. VII Congreso Iberoamericano de Nefrología y XLII Congreso Nacional de la Sociedad Española de Nefrología. Gran Canaria. October 2012.

[2] Cook D et al. Internet-based learning in the health professions: a metaanalysis. JAMA. 2008;300:1181–1196.

[3] Lam-Antoniades M, Ratnapalan S, Tait G. Electronic continuing education in the health professions: an update on evidence from RCTs. J Contin Educ Health Prof. 2009;29:44–51.

[4] Cook DA, Levinson AJ, Garside S, Dupras D, Erwin PJ, Montori V. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Academic Medicine. 2010;85:909–922.

[5] Harris, J. M., Sklar, B. M., Amend, R. W. and Novalis-Marine, C. (2010), The growth, characteristics, and future of online CME. J. Contin. Educ. Health Prof., 30: 3–10.

[6] Liyanagunawardena TR, Williams SA. Massive Open Online Courses on Health and Medicine: Review. J Med Internet Res 2014;16(8):e191.

[7] Llambí L, Margolis A, Toews J, Dapueto J, Esteves E, Martínez E, Forster T, López Arredondo A, Lockyer J. Distance education for physicians: adaptation of a Canadian experience to Uruguay. J Contin Educ Health Prof. 2008 Spring; 28 (2):79-85.

[8] Llambí L, Esteves E, Martinez E, Forster T, García S, Miranda N, Lopez Arredondo A, Margolis A. Teaching tobacco cessation skills to Uruguayan physicians using information and communication technologies. J Contin Educ Health Prof. 2011 Winter; 31(1):43-48.

[9] Martínez E, Margolis A, Joglar F, Alonso D, López Arredondo A, García S. From Uruguay to Puerto Rico: similarities and differences in CME and the context of health care. CME Congress, Vancouver, Canada, May 2008.

[10] Margolis A, Santiago B, Martínez E, Dapueto J, Lorier L, Fernández Z, Forster T, Miranda N, Muñoz C y López A. Four-year Collaboration in CME Among Latin American Institutions. CME Congress, Toronto, Canada, May 2012.

[11]Cohen H, Margolis A, González N, Martínez E, Sanguinetti A, García S, López A. Implementation and evaluation of a blended learning course on gastroesophageal reflux disease for physicians in Latin America. Gastroenterol Hepatol. 2014 Aug-Sep;37(7):402-7.

[12] Medeiros N, Pérez G, Sniadover N, López A. Generador de casos interactivos. Presented at MoodleMoot Uruguay 2011, available at: http://www.moodlemoot.org.uy/moodlemoot_2011/moodlemoot/moodlemootuy2011_submission_26.pdf

[13] Mansouri M, Lockyer J. A meta-analysis of continuing medical education effectiveness. J Contin Educ Health Prof. 2007 Winter;27(1):6-15.

[14] Olson, C. A. (2012), Twenty predictions for the future of CPD: Implications of the shift from the update model to improving clinical practice. J. Contin. Educ. Health Prof., 32: 151–152.

[15] Fletcher SW. Chairman's Summary of the Conference. In: Hager M, editor. Continuing Education in the Health Professions: Improving Healthcare through Lifelong Learning; 2007 Nov 28 – Dec 1; Bermuda. New York: Josiah Macy, Jr. Foundation; 2008.).

[16] Maloney S, Haas R, Keating JL, Molloy E, Jolly B, Sims J, Morgan P, Haines T. Breakeven, cost benefit, cost effectiveness, and willingness to pay for web-based versus face-to-face education delivery for health professionals. J Med Internet Res. 2012 Apr 2;14(2):e47.

[17] Margolis A, Parboosingh JT. Networked learning and network science: potential applications to health professionals' continuing education and development. Accepted for publication at J Contin Educ Health Prof.

Address for correspondence

Alvaro Margolis, MD MS

Email: alvaro.margolis@evimed.net

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-376

Danish Citizens and General Practitioners' Use of ICT for their Mutual Communication

Pernille Bertelsen[a], Lone Stub Petersen[a]

[a] *Danish Centre for Health Informatics and Department of Development and Planning, Aalborg University, Denmark*

Abstract

This paper reports on selected findings from a Danish national survey of citizens' perception and use of information and communication technology (ICT) for their health care [1]. Focus is on citizens' use of ICT and on communication with their General Practitioner (GP). It also focuses on citizens' experience of their GPs' ICT use and no use during medical consultations. The responsibility for medical service in Denmark is to a large extent handed over to the primary sector where the GP is the gatekeeper. Our data display that 65% of the adult citizens or their relatives have been using ICT to communicate with their GP. Twenty-two percent have experienced their GP use a computer screen to actively show them something while they have a consultation. Further, our data supports the assumption that the higher the education people have, the more likely they are to use ICT for their health care. The understanding of the use of ICT in communication with the GP is central to monitoring and developing an ICT that supports all citizens and considers new ways in which to enhance quality of care.

Keywords:

Citizens; General Practitioner; eHealth.

Introduction

Citizens' use of Information and Communication Technology (ICT) for their health care (eHealth) has received increased attention in Denmark. The public health care system is under pressure (demography, chronic illness, comorbidity, etc.). eHealth has, to a large extent, given better quality and also a faster turnover of patients in hospitals. In Denmark, the number of days a patient is admitted to a hospital has decreased from an average of 9.8 days in 1978, to five days in 2002, to 3.6 days in 2013 [2][3]. From 2005 to 2010, the number of outpatient visits increased by 32% [4]. The Danish health care sector is witnessing a structural change; responsibility for medical service to an even larger extent is being allocated to the primary sector where the General Practitioner is the main gatekeeper. Investments in advanced health care technology including eHealth make the hospitals provide treatment that is more specialised, leaving the primary sector (Municipalities and General Practitioners) to be responsible for care and medical follow-up.

Denmark has an international reputation of being a country of fast adopters of eHealth [5]. An international survey in the 2010 Commonwealth Fund International Health Policy Survey found that <10% of adults in the 11 countries that participated in the survey had used email to pose a medical question to their regular GP or place of care in the last two years [6]. In contrast, the 2013 survey (n=1,058) of Danish citizens' use of eHealth, from which we report in this paper, disclosed that 65% of respondents, or their closest relatives had used ICT to communicate with their GP [1].

ICT has the potential to play a pivotal role in facilitating contact between citizens' and their health care providers in daily health care. In the communication between GPs and citizens there has been a slow but steady change in communication as seen in Table 1. From 2007 to 2013 there has been a steady increase in e-mail consultations from 0.8 million to 4 million while phone consultations have decreased from 14.7 million to 11.8 million [7]. The reasons for this change are many; e.g., structural changes of payment for GP consultation, changes within GP clinics like cutbacks on telephone hours, the availability of e-mail consultation as well as the citizen's level of ICT literacy [8].

Table 1 – Development in citizen's communication channels with GP's [7].

	2007	**2010**	**2013**
Face to face consultations (daytime)	18.580.394	19.112.781	19.951.578
GP telephone consultations (daytime)	14.778.738	13.834.199	11.816.462
Email consultations	802.581	2.266.020	4.073.007

The GPs are gatekeepers in the Danish healthcare system and an increasing amount of tasks are moved to the primary sector. Understanding the use of ICT in communications between citizens and the GP becomes central to the development of Danish health care services and important in relation to changing and/or expanding the possibilities of ICT use in this area. Further, it is important to have a focus on the citizens' perspective. For many stakeholders within Danish healthcare, the central role of the citizens in eHealth is debated; healthcare managers, healthcare professionals, citizens, patient associations, system developers, vendors, and politicians advocate that eHealth has the potential to enhance citizens' positions in treatment and care [8][9]. The 2011 and 2013-17 Danish National and Regional eHealth strategies support the potential of using eHealth as a means to involve and empower the citizens in healthcare [8][9]. In the 2014 government health care policy and investments strategy for health and

welfare it is expressed as: "Patients across the country must take advantage of the new technological possibilities for telemedicine treatment and follow-up at home, where it provides better and cheaper treatment" [10]. However, it is not enough to focus on providing more efficient and effective treatment. In order to support patient empowerment, we need to understand the citizens perspective [11] and make sure that we avoid developing ICT for "People like us", that is for well educated, ICT-literate system developers and health care professionals [12].

In 2013, inspired by Canadian and Australian studies of consumer experience with eHealth [13][14], the Danish Center for Health Informatics (DaCHI) commissioned MEGAFON, a Danish market research agency, to carry out the first National Survey on Danish citizens expectations and perspectives on eHealth with a population sample of (n=1,058) [1]. DaCHI has many years of experience in monitoring eHealth implementation in Denmark, e.g., the national implementation of Electronic Health Record (EHR), as well as more recent experience with the national monitoring of health care professionals' use of health informatics in their daily clinical practices [15]. This paper reports on selected findings from the survey [1]. The following questions were investigated:

- Have you or your closest relative used ICT to communicate with your GP? (Yes/No)
  - (Yes) For what purpose did you use ICT to communicate with your GP?
  - (No) Why do you or your closest relative not use ICT to communicate with your GP?
- Last time you visited the GP clinic; did your GP use a computer while you were present? (Yes/No)
  - (Yes) Did the GP use the computer screen to show you anything?

Method

The study questionnaire was drafted with inspiration from the Canadian consumer survey [13]. The baseline data consists of the number of children in the household, educational level, access to Internet (both in private and at work/school), mobile phone and smartphone, and chronic disease occurrence. For those questions inquiring about citizens' views and attitudes, it was possible to elaborate on the answers.

We tested the questionnaire twice in minor pilot studies both asking the respondents to fill in the questionnaire and give feedback on the design of the questions, as well as adding the issues they experienced and elaborating on their views. Then, the refined survey was handed over to the market research agency. They tested the survey again, and certain questions were reformulated. The enumeration was designed as a combinational survey where both e-mail and telephone were used to question respondents from their citizen panel. In total, the survey had 1,058 respondents, which is equivalent to the needed number of respondents for a country which is the size of Denmark. Of the respondents, 80% participated by e-mail and 20% by telephone.

In addition to general baseline data and subjects concerning the use and perception of eHealth, the respondents were questioned whether they or their closest relative had used ICT to communicate with their GP. If that was the case, they were further asked what issues they had communicated about.

Further, we were also interested in getting an insight into citizens exposure to use of ICT when they physically visited the GP's office. Therefore, we asked the respondents if they had experienced their GP using a computer while they were present in the consultation room? And additional, if the GP used the computer screen to show them anything during consultation.

The population comprised of people who were already part of a citizen panel. Two-thousand one-hundred e-mails were sent, and 500 phone numbers were called. The selected respondents reflected the Danish adult population with respect to age, education and geographic distribution. Of the 1,058 respondents, 51% were female, and 49% were male. The population age was between 18 and 70+ years, with a distribution of 15% to 19% of the respondents in the six age groups.

The data was sampled over a period of one week in October 2013, and sorted into frequency and cross tables for further analysis.

Results

As already mentioned in the introduction 65% (n=684) of respondents reported that they or their closest relatives had communicated with their GP, using ICT while 33% (n=350) had no experience in doing so. The difference between age groups was minor. Of respondents between 18 and 29 years old, 68% of them or their relatives had used ICT to communicate with their GP; in the age group 70+ years, it was 62%. Seventy-one percent of the female respondents had used ICT and 59% of the male.

When we looked at the educational background of the respondents', 72% of those holding a medium or higher educational background or their relatives, had been using ICT, while 46% of those with no professional education or their relatives had used ICT in connection with their healthcare.

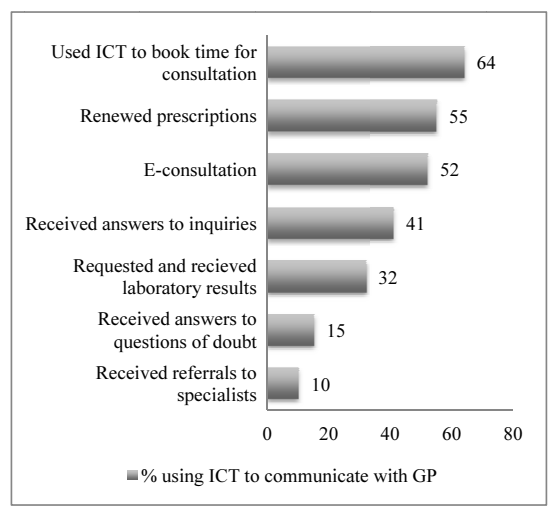


Figure 1 – Answers to questions about the use of ICT in communication with the GP (n=684)

The respondents that had experience using ICT to communicate with their GP were asked to elaborate on the activity or activities for which they had used ICT (Figure 1). Out of the 65%, a majority (64%) of the citizens had used ICT to book a consultation, while 55% had renewed their prescriptions. Fifty-two percent had used e-consultation, and 41 % had received answers to inquiries, while 32 % had been requesting for laboratory and blood test results using ICT. Fifteen percent received answers to questions of doubts, while 10% received referrals to specialists.

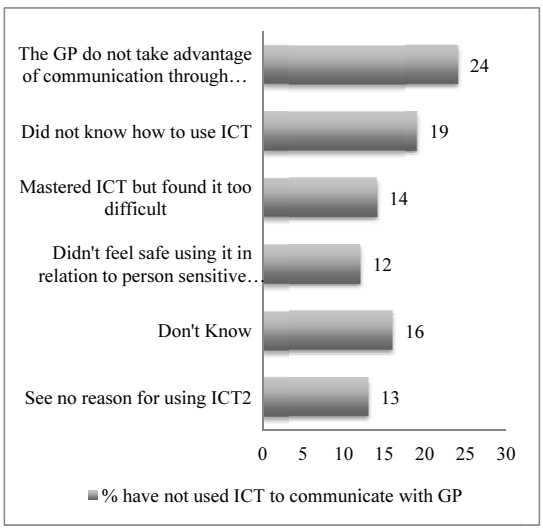


Figure 2 – Answers to questions about not using ICT in communication with the GP(n=350)

Figure 2 concentrates on the 33% who had not used ICT for communication with their GP. Twenty-four percent stated the primary reason to be that their GP did not encourage communication through ICT. Nineteen percent stated that they did not know how to use ICT, while 14% expressed that though they master ICT, they find it too difficult. Further, 12% did not feel comfortable using ICT for communicating about person-sensitive issues, while 16% answered, "Do not know." Twenty-six percent selected "other" as a reason for not making use of ICT to communicate with their GP. When they were asked to elaborate, 13% stated that they saw no need to use ICT to communicate with their GP. Other respondents indicated that they preferred to be physically present during consultation, while others commented that they preferred to use a telephone to contact their GP.

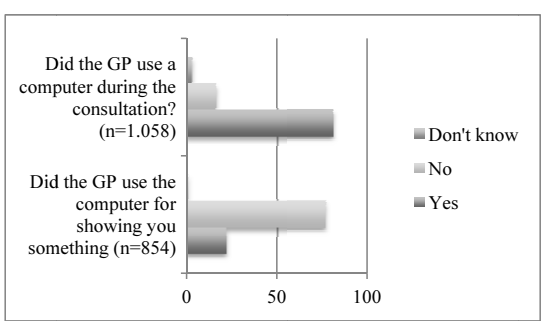


Figure 3 – The GP's use of ICT to communicate with the Citizens

The data in Figure 3 show that the majority of Danish GPs use computers during a consultation while the citizens are in the consultation room. Eighty-one percent (n= 854) of the respondents had that experience, while 16% (n=170) had not.

Twenty-two percent of the respondents (n=99) experienced that ICT was used interactively by the GP to show them something on the screen during consultation. There were differences between the age groups. Of respondents between 18 and 29 years old, 18% had been shown something during consultation, while this was the case for 36% from the age group 70+.

Of the respondents, 27% with a long professional educational background, and 20% without any education had that experience during a consultation with their GP.

Discussion

The Danish health care sector is undergoing structural changes and the important role of the primary sector is being emphasized. These changes make it pivotal that citizens can have easy access to their GP as he/she is their primary contact and gatekeeper to other health care services.

The activities that had the highest score were to book appointments, renew prescriptions, ask health related questions in e-consultations and get information on test results. More or less the same issues that the telephone consultation with the GPs clinic also can provide answers to. However, the number of email contacts is much higher than the decrease in telephone contacts. The difference from using the telephone is the flexibility, making communication independent of time and place. More women than men are using ICT to communicate with their GP. Data from the National Danish Statistic Service show that women generally have greater contact with the GPs than men. In 2007 men made 5.4 million phone calls to their GP and sent 0.3 million e-mails while women made 9.3 million calls and sent 0.5 million e-mails. In 2013, men made 4.4 million calls and sent 1.3 million e-mails, while women made 7.4 million calls and sent 2.6 million e-mails [7]. The difference in contact with the GP can partly be explained by maternity related consultations, however further studies would be needed to verify this hypothesis.

The General Practise is allowed a flexibility in organizing their work when the citizens use ICT to communicate with the GP. They do not have to answer the phone and deal with citizens needs in real-time but can attend them whenever time allows. The citizens on their part get the flexibility by not having to meet telephone hours but can write and request for information and services 24/7.

When we look at the educational background of respondents, 72% of those holding a medium or higher educational background, had been using ICT, while 46% of those with no professional education had used ICT to communicate with their GP. Therefore, our data support the assumption that the higher the education, the more likely the citizens are to use ICT for contact with their GP. Knowing that, in Denmark, the severe healthcare problems, followed by high public expenditures, are to be found among citizens with no or very low educational background (and, hence, low socioeconomic status) [16], it should be taken seriously that the citizens with these attributes are weakly represented in use of ICT. As mentioned in the introduction, a solution to this could be to focus development of eHealth on the needs of the low educated citizens perspective, not just on the perspective of "people like us" [11][12].

Eighty-one percent of the citizens had experienced GPs using ICT during a consultation. This number indicates that ICT is aspiring to become a well-integrated part of the GPs clinical work practise and their communication with patients in the consulting room. However, our data does not specify the different tasks for which the GPs use ICT, except one. The fact that 22% of the respondents had experienced the GP show them something on the computer screen, is an indication that ICT is being used during consultation. It demands further studies to assess the full status and potential for integration and use of ICT in consultations between GP and citizens. Data further indicated that elderly citizens had a slightly higher experience of the GP showing them something on the

computer screen which could indicate that they are more often consulting the GP and with more complex issues than younger people. For citizens with a high education, the percent that had experienced the GP using a computer screen were slightly higher than for those with no education. This could be an indication that this group asks questions and pays attention to understand how and why. However, the difference is not big enough to make any conclusions without further investigation.

Conclusion

According to the pressented self-assesment survey of Danish citizens, more than half of adult citizens or their closest relatives use ICT to communicate with their GP. ICT is used by citizens of all age groups and educational background though not equally distributed. The increase in use of ICT in communication with GP is also evident in the increase in e-mail consultations from 0.8 million in 2007 to 4 million in 2013. However, the survey also showed a difference in use of ICT between men and women, and between different age groups. There can be different reasons for this, but evaluation and further studies are needed in order to understand and possibly bridging the social gaps in use of ICT for health care.

The survey showed that the higher the education level, the more likely citizens are to adapt to new eHealth services. This raises the concern that eHealth is designed by and for people with high education ("people like us") [12]. Designing eHealth for people like us affects other social groups' ablility to access and use eHealth. People with low education and low eHealth literacy but large health problems (the disempowered, disengaged and disconnected – DDD's) need to participate in defining and designing eHealth that supports their needs and comptencies [12].

As more and more heath care services, telemedicine and eHealth is targeting GPs in the primary health care sector, it becomes vital to pay attention to design *with* citizens that have the largest health problems and low skills to support their own treatment and care, and not *for* these social groups. The adoption of ICT is increasing, however, we should not forget that with a public health care system we should be designing eHealth for all, especially those with the most need.

References

[1] Tornbjerg K, and Bertelsen P. Undersøgelse af borgernes anvendelse af sundheds-it i 2013 – en udforskning af danskernes kendskab, holdninger og forhold til it, til gavn for eget helbred, Aalborg Universitet, Dansk Center for Sundhedsinformatik. Marts 2014.

[2] Sundhedsstyrelsen. Landspatientregisteret gennem 25 år 1977-2002. Sundhedsstyrelsen, 2003: 21: p. 5.

[3] Statens seruminstitut. Sundhedsdata på regions-, hospitals- og kommunalt niveau, www.esundhed.dk, 2014.

[4] Ministry of Health. Fakta ark, om omlægning fra stationær til ambulant behandling, Ministry of Health, Denmark, 2012.

[5] Bhanoo SN. Denmark leads the way in digital care. New York Times, 2010.

[6] Commonwealth Fund International. Health Policy Survey, 2010, *www.commonwealthfund.org,* Chartpack full, slide 23, 2010.

[7] Danmarks Statestik. SYGK: Lægebesøg mv. med offentlig tilskud efter område, ydelsesart, alder, køn og socioøkonomisk status 2006-2013. Danmarks Statestik, 2015.

[8] Jakobsen KT. Øget brug af email-konsultationer hos de praktiserende læger har betydet højere udgifter. Fonden Kraka, 11. June 2013.Statens Serumsinstitut. Digitalisering af sundhedsvæsenet 2013–2017, Statens Serumsinstitut, Juni 2013.

[9] Danske Regioner. Strategi for IT-understøttelse af patient empowerment, for et trygt, sikkert og effektivt sundhedsvæsen, RSI, Danske Regioner, Marts 2011.

[10] Regeringen. Jo før – jo bedre: Tidlig diagnose, bedre behandling og flere gode leveår for alle. Regeringen, Denmark, August 2014.

[11] Petersen LS, Bertelsen P. Citizen personas: exploring challenges of citizen-centric eHealth. MIE2015, Stud Health Technol Inform 2015; 210: 582-586.

[12] Showell C, Turner P. Personal health records are designed for people like us. MedInfo 2013; pp. 1037.

[13] Zelmer J. Health information solutions for consumers: research and plans. Healthcare Management Forum 2011; 24 (3): 144-146.

[14] Consumers Health Forum of Australia. eHealth and Electronic Health Records. Current Consumer Research 2011.

[15] Tornbjerg K, and Nøhr C. Undersøgelse af klinisk anvendelse af sundheds-it-systemer 2013, Aalborg Universitet, Dansk Center for Sundhedsinformatik. Marts 2014.

[16] Mackenbach JP, Stirbu I, Roskam AR, Schaap MM, Menvielle G, Leinsalu M, Kunst AE. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008; 358 (23): 2468-2481.

Address for correspondence

Pernille Bertelsen: pernille@plan.aau.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-380

The Online Availability of Multilingual Health Promotion Materials Produced by Local Health Departments: an Information Assessment

Daniel Capurro[a,b], Shomir Chaudhuri[b], Anne M. Turner[b,c]

[a] *Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[b] *Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA*
[c] *Department of Health Services, School of Public Health, University of Washington, Seattle, WA, USA*

Abstract

Objectives*: Local Health Departments (LHDs) are a key source of health promotion information. For ethnically and culturally diverse communities, it becomes important to provide minorities with language appropriate health information. This project sought to assess the availability of multilingual health promotion materials on LHD websites in Washington State (WA), USA.* ***Methods****: We performed a cross-sectional study of all 34 LHD websites in WA. We collected and classified health promotion documents available to the public, specifically, whether translated versions were available. We also assessed the extent of document sharing between LHDs.* ***Results****: We identified 1,624 documents across 34 LHDs. Topics most frequently covered were communicable diseases and emergency preparedness. Fewer than 10% of documents were available in non-English languages. We found little evidence of document sharing between LHDs; only 5% of all documents were shared between LHDs.* ***Conclusions****: WA LHDs provide a variety of health promotion materials for the public, but few multilingual materials are available online. New technologies for facilitating document sharing and machine translation may improve the present landscape.*

Keywords:

Public health practice; Translation; Health Promotion; Limited English Proficiency; Public Health Informatics.

Introduction

Washington State (WA), like the rest of the U.S., has become increasingly culturally and linguistically diverse. In the past decade, the Hispanic and Asian populations in WA have increased by 71% and 49%, respectively [1, 2]. Many of these individuals do not speak English well and are considered to be limited English proficient (LEP) [3]. Individuals who are LEP have less access to health care and worse health outcomes [4-7]. Low English literacy has been consistently associated with reduced access to recommended health care services [8] and increased prevalence of risk factors for chronic health conditions [9], a sign of currently existing health care disparities. Reducing such disparities is a U.S. priority [10]. Implementing culturally and linguistically appropriate services in health care is one of the means by which Healthy People 2020 [11] intends to address these disparities.

The need for culturally and linguistically appropriate written materials is well acknowledged. However, the availability of high-quality health information in languages other than English is limited. Health information translated into Spanish, the second most common language in the U.S., has been shown to be of poor quality and inconsistent [12,13]. This limited availability of health information materials in non-English languages in the U.S. persists despite legal requirements to provide equal access to health services for LEP individuals [14].

Public health agencies have a very important role in educating the population about issues of public health importance. One of the ten essential public health services is to "inform, educate and empower people about health issues"[15]. At the local level, this includes providing "health information, health education, and health promotion activities to reduce risk and promote better health" [15]. This makes LHDs, the fundamental territorial unit enabled to provide public health services in the U.S., an ideal avenue for disseminating culturally and linguistically appropriate health promotion materials to LEP communities. Furthermore, people are increasingly obtaining health information online. The 2007 Health Information National Trends Survey (HINTS) conducted by the National Cancer Institute (NCI) found that 58% of the U.S. population obtains health information through the Internet, up from 50% two years before [16].

Despite the importance of the Internet for disseminating health information, and the increasing frequency with which consumers are using such information, little is known about the current state of health promotion information provided by LHDs through the Internet, and the availability of multilingual health promotion materials for LEP individuals. To our knowledge, only one study has addressed this question [17].This study surveyed public and private mental health care providers in New Mexico about their current practices around language access services, including the availability of translated materials. The authors reported that 81% of agencies translated at least one piece of consumer education material, with a higher percentage of translated materials in locations with a higher proportion of non-English speaking residents. Although the study offers interesting insights regarding language access practices, the investigation was limited to mental health care organizations, relied solely on self-reported information, and the coverage of translated materials was not quantified.

Our own research investigating the potential use of machine translation in public health settings indicated that time and costs were the major barriers for providing multilingual health materials [18]. Since machine translation relies on the availability of training materials—that is, pairs of documents in two languages—the availability of manually translated documents in public health promotion was the main goal of this study. In addition, we sought to better understand the availability of such documents and to assess whether LHDs were meeting the LEP health information needs of their communities. In particular, we aimed to identify the range of topics covered, the types of health education materials, and the percentage of translated materials available online. We investigated the correlation of available translated materials with the percentage of non-English speaking residents in each county. Finally, we describe the use of a specific public health terminology to classify the documents we identified.

Materials and Methods

Study Setting and Document Identification

We identified all 39 counties and the 34 associated LHDs in Washington State. For each LHD we located the corresponding website. We then systematically reviewed each website to identify English and non-English health promotion materials over a five-month period. Health promotion materials were defined as (1) written documents targeting the general population that (2) provided information on how to prevent, manage or treat health conditions, or modify risk factors of public health relevance. We included all documents hosted on the LHD's website produced by the LHD. We excluded documents produced by other entities and links to externally hosted documents, forms and documents that only provided contact information to public health services and programs, and those targeting professional groups, health providers or other public health workers.

Document Classification

For each document, we captured the title, the topic, and the availability of a translated version of the document. In order to estimate the overall translation burden handled by LHDs, we also classified documents based on the amount of text. Each document was classified into three categories: (a) Low Text, for documents with mostly images and a few words or sentences and no complete paragraphs; (b) Medium Text, for documents with at least one complete paragraph and up to two pages in length; (c) High Text, for text-heavy documents of more than two pages in length.

To identify a suitable terminology to classify the topics covered by the health promotion documents, we conducted a thorough literature search using Medline (www.pubmed.gov), the National Centers for Biomedical Computing BioPortal and the Unified Medical Language System (UMLS) (www.nlm.nih.gov/research/umls/). The terminology had to include health promotion or public health topics and be comprehensive enough to describe the full range of available public health documents. Despite the importance of health promotion, we were unable to find a standard terminology on health promotion or public health topics that fulfilled all our requirements. The terminology that best approximated our needs was the European Multilingual Thesaurus on Health Promotion (EMTHP) [19]. We adopted this terminology and extended it where necessary. This included adding new terms and modifying existing terms to reflect the current topic classification practices used by the LHDs we studied.

Statistical Analysis

We conducted descriptive statistics to quantitatively describe the current status of the availability of multilingual health promotion materials. We also calculated the correlation coefficient between the proportions of multilingual documents available and non-English speakers in every LHD.

Results

Document Collection

We identified websites for the 34 LHDs serving WA's 39 counties (three LHDs serve more than one county). All LHD websites contained written health promotion materials. The number of documents available on each website varied greatly and did not have a normal distribution. The median was 24 documents per LHD website, with a range of 2 to 289 health promotion materials available on a given website. Public Health – Seattle & King County (PHSKC), the largest LHD in the state, offered the largest number of health promotion materials. In total, we identified 1,624 English health promotion documents from the 34 LHD websites.

The five topics most frequently covered were Communicable Diseases (36%), Environmental Health (22%), Emergency Preparedness (7%), Addictions (4%) and Injury Prevention (3%). In addition, 8% of the identified documents were newsletters covering more than one topic. Coverage for chronic and non-communicable diseases was remarkably low; only 1.4% of all documents covered these highly relevant conditions. The vast majority of documents (88%) were under two pages long. Longer documents (> 2 pages) were primarily newsletters, detailed handouts and small booklets (12%).

Translated Materials

Of the 1,624 identified documents, 152 (9.4%) were translated into at least one additional language. All translations were available in Spanish. Among the other available languages, most frequently are Vietnamese, Chinese, Korean, Russian, and Somali. The majority of non-Spanish translations came from PHSKC.

The most frequently translated topics were Cancer (57%), Geriatric Health (50%) and Mental Health (25%); however, the absolute number of English language documents available for these topics was remarkably small, with 7, 2 and 8 documents available respectively. Of the most frequently covered topics, 13% of Communicable Disease documents were translated; 5% of Environmental Health and 11% of Emergency Preparedness documents had translations available. No public health newsletter was translated into any non-English language (Table 1).

Table 1 - Frequency of translated health promotion materials for the most commonly available topics on Washington State Local Health Department websites

Topic	Translated – n (%)	Not Translated – n (%)	Total
Communicable Diseases	76 (13.1%)	502 (86.9%)	578
Environmental Health	18 (5.1%)	333 (94.9%)	351
Emergency Preparedness	15 (11.5%)	115 (88.5%)	130
Injury Prevention	4 (7.3%)	51 (92.7%)	55
Addiction	8 (13.8%)	50 (86.2%)	58
Chronic Diseases	0 (0%)	23 (100%)	23

The percentage of translated documents varied widely across different LHDs. The median percentage of translated documents was 2% and ranged from 0% to 50% (Adams County); 15 of the 35 LHDs (43%) had no translated materials on their websites.

The proportion of translated health promotion documents was highly correlated with the proportion of non-English speaking individuals living in each county, $r(30) = 0.90$, $p < .01$. This indicates that counties with higher proportion of non-English speakers are translating a greater proportion of their health promotion materials. Despite this positive correlation, several LHDs (such as Chelan and Douglas) that serve counties with a high proportion of non-English speakers offered remarkably few translated health promotion materials (Figure 1).

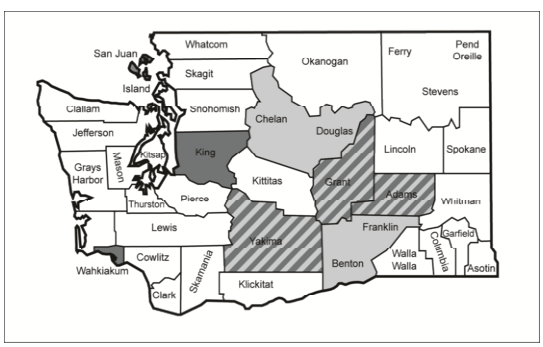


Figure 1 - Map representing Local Health Department (LHD) jurisdictions with the highest proportion of non-English speakers (light gray) and the highest proportion of translated documents (dark gray). Striped areas represent LHD jurisdictions where these two conditions overlap

Document length restrictions

We also conducted an analysis of documents produced by entities outside each LHD. We identified an additional set of 380 documents, 19% of the total documents hosted on LHDs' websites. When looking at the original source of these documents, we found little evidence of document sharing between LHDs. The most frequent external sources were the Washington State Department of Health (WA DOH), with 77 documents (20%), and the CDC with 62 (16%). Only five percent of the documents were created by other LHDs in Washington State, with the most frequent source being PHSKC.

We found little evidence of sharing of translated documents between LHDs. Twenty percent of all translations available in the state were authored by another department, similar to the proportion of shared English materials. Only one document was produced by a different LHD within Washington State; either the WA DOH or departments outside the state had originally produced the rest.

Terminology

Despite the fact that the EMTHP was specifically developed to cover health promotion materials, we found it was not sufficiently broad to adequately classify many of the health promotion documents collected in our study. Although we were able to classify most documents using this terminology, many of the principal categories listed in Table 1 were not included in this thesaurus. One notable example was the insufficient representation of environmental health topics and the complete lack of an "Emergency Preparedness" category. The EMTHP had a "Disasters" category with very few sub-concepts, which did not capture the large number of documents produced by LHDs for informing the public about Emergency Preparedness. In addition, the EMTHP did not have a general "Chronic Disease" category. On the other hand, the variety and number of available communicable disease topics was insufficient. For example, extremely relevant public health communicable diseases such as varicella and pertussis were not classified under the "Communicable Diseases" category but were assigned to "Childhood Diseases." We chose to classify all communicable diseases of public health importance into a broader "Communicable Diseases" category. Overall, our impression was that the EMTHP was "clinical condition centric", as opposed to "health problem or program centric," which is how most Washington State LHDs organize topics.

Discussion

Increasingly, the Internet is becoming a crucial method for distributing information to the public [20]. This includes dissemination of health information to providers and the general population by government and other health care organizations [22, 22]. In addition, seeking health information on the Internet has become common. This provides a unique venue for delivering high-quality public health information for vulnerable communities. However, the accessibility and quality of language-appropriate health information on the Internet remains a concern, especially for LEP individuals.

We found more than 1,600 health promotion documents available on LHD websites. Those documents were available for the general public and for public health practitioners to print and distribute. Our finding that fewer than 10% of all WA LHDs websites' health promotion documents are available in non-English languages underscores the lack of availability of translated materials. This situation is concordant with previous findings. A 2001 study found that a majority of websites only covered clinical topics minimally, and all English "and 86% of Spanish websites required high school or greater reading abilities" [23]. Left unchecked, this situation may further widen health disparities for LEP communities.

Multiple reasons could account for the lack of non-English documents in spite of federal and state requirements to provide multilingual materials. Although outside the scope of this study, informal interviews suggest that the high cost of current translation processes is a reasonable explanation. Numerous steps are involved in producing a culturally and linguistically appropriate translation of a health-related document [24], which ultimately affects the overall cost of the process. The burden of manually translating documents and the associated costs is probably not sustainable for underfunded departments such as LHDs, thus explaining the small number of translated documents. Additional forms of disseminating culturally and linguistically appropriate documents, such as video or audio documents are even more expensive to produce.

Informatics tools and principles can aid public health in fulfilling the need for translated health materials in several ways. One way is through web-based document sharing systems. We conducted an additional analysis of documents that were produced by other LHDs to estimate the amount of document sharing. In this case, we found only one translated document that had been produced by another LHD in Washington State. Overall, external entities produced only 20% of all translated materials. This is consistent with findings of a statewide survey, also conducted by our group, which showed that almost 80% of LHDs in the state did not share their translations or did not know whether efforts to share translations existed [25]. There is definitely room for improvement. This is especially relevant in the context of previous efforts implemented by Washington State.

The WA DOH has established the Health Education Resource Exchange (H.E.R.E) (http://here.doh.wa.gov/). This is a website that hosts, among other content, health education materials for use by other LHDs for free. Some of these materials are translated into multiple languages. Despite the availability of this trusted resource, and the number of translated documents it contains, its contribution to the overall pool of translated documents in our study was small. In our experience, LHDs tend to tailor documents to their communities, including local information and context.

Another approach to increasing the availability of translated health promotion documents is to partially automate the translation processes though statistical machine translation (SMT). Google Translate is an example of a free and publicly available SMT (http://translate.google.com). At least one health department currently uses SMT for translating their English materials into Spanish [26]. Although such tools hold promise for making translated materials more readily available, they currently perform poorly in the domain of public health [27]. SMT optimized for health promotion documents has the potential to reduce the time and costs associated with human translation of documents and increase their availability to LEP communities [27, 28]. The use of SMT for public health information has been insufficiently explored. Our research group will investigate the use of SMT to improve access to public health information. However, the small proportion of translated materials may severely limit the ability to adequately train a statistical machine translation system.

At the beginning of our study, we chose the EMTHP as the terminology to classify the documents. Although not surprising, it was remarkable to find the mismatch between the way LHDs organize their documents and the way the EMTHP is built. LHDs organized their documents based on the public health programs and services they offer. In contrast, the EMTHP has a clinical focus, which fails to capture this feature of US public health. As with other clinical vocabularies, the EMTHP is centered on classifying by disease conditions without considering public health importance. The current lack of a terminology to adequately classify public health documents is a significant gap. Despite all the progress that has been made in medical terminologies and ontologies, public health remains underrepresented ontologically. A 1997 study found that terminologies focused on clinical information systems, and only 8% of terms "were specifically identified as needed for population-based public health" [29]. Thirteen years later, preliminary results of a project to extend the Unified Medical Language System to include public health were presented at the 2010 Annual Symposium of the American Medical Informatics Association [30] but final results are still unavailable. Further work on creating public health domain specific terminologies or modifying existing terminologies to reflect the domain of public health are needed to improve the organization and access of health materials for all residents, including those who are LEP.

Limitations

We only included documents that were posted online, which means that we could have missed translated documents available only in print. However, in a recent survey conducted by our research team, with a 59% response rate from LHDs in Washington State, 86% of LHDs reported using the Internet to communicate non-urgent information to their communities [25], which suggests that the impact of only looking at materials online was minimal. We also limited our study to Washington State LHDs. Although the programs and services provided by these health departments are typical of other LHDs across the country, it is possible that the characteristics and the availability of translated materials may differ from state to state. However, the variety of size and complexity of the LHDs included might still represent realities outside this state. Finally, we focused our document collection on LHDs and did not include other relevant public health organizations such as non-governmental organizations (NGOs) or large integrated health maintenance organizations (HMOs), which are important sources of consumer-based health promotion materials.

Conclusions

Despite U.S. federal and state requirements that public health departments make available such materials, the number and scope of translated materials on Washington State LHDs' websites is limited, a situation that might not change easily if we consider the burden of translating health promotion materials. Potential application of new information technologies, such as document tools that help foster document sharing and exchange practices between LHDs, and automatizing aspects of translations with statistical machine translation should be investigated for their potential in improving access to low cost, high quality, bilingual health promotion materials. The limited availability of publicly available manually translated public health promotion documents might limit the ability to train such a system.

Acknowledgements

The authors wish to thank Megumu Brownstein for her assistance with document collection and classification, Kate Cole and Carmen Kwong for their assistance in manuscript preparation, and Daisye Orr for providing information

regarding the H.E.R.E. project. The research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM010811 and the Fulbright-MECESUP Faculty Development Scholarship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] US Census Bureau. (2010). Washington QuickFacts. Retrieved from http://quickfacts.census.gov/qfd/states/53000.html

[2] US Census Bureau. (2000). American FactFinder: Washington. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmkAmerican

[3] Community Survey Reports (ACS). (2011). Language use in the United States. U.S. Census Bureau. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_11_1YR_DP02&prodType=table

[4] Phillips KA, Mayer ML, Aday LA. Barriers to care among racial/ethnic groups under managed care. Health Affairs 2000; 19(4): 65-75.

[5] Agency for Healthcare Research and Quality (AHRQ). (2006). National Healthcare Disparities Report. Retrieved from http://www.ahrq.gov/qual/nhdr06/nhdr06.htm

[6] Fiscella K, Franks P, Doescher MP, Saver BG. Disparities in health care by race, ethnicity, and language among the insured: findings from a national sample. Medical Care 2002; 40(1): 52-59.

[7] Beal AC. (2009). Health Care Disparities - The Commonwealth Fund. Health Care Disparities. Retrieved from http://www.commonwealthfund.org/Content/Program-Areas/Archived-Programs/Health-Care-Disparities.aspx

[8] Cheng EM, Chen A, Cunningham W. Primary language and receipt of recommended health care among Hispanics in the United States. Journal of General Internal Medicine 2007; 22(Suppl 2): 283-288.

[9] DuBArd CA, Gizlice Z. Language spoken and differences in health status, access to care, and receipt of pre-ventive services among US Hispanics. American Journal of Public Health 2008; 98(11): 2021-2028.

[10] US Department of Health and Human Services (DHHS). (2011). HHS Action Plan to Reduce Racial and Ethnic Disparities: A Nation Free of Disparities in Health and Health Care. Washington, DC.

[11] US Department of Health and Human Services (DHHS). (2010). Healthy People 2020. Washington, DC: Office of Disease Prevention and Health Promotion.

[12] Berland GK. Elliott MN, Morales LS. (2001). Health information on the Internet: Accessibility, quality, and readability in English and Spanish. Journal of American Medical Association 2001; 285(20): 2612-2621.

[13] Turner A M, Stavri Z, Revere D, Altamore R. From the ground up: Determining the information needs and uses of local public health nurses in Oregon. Journal of the Medical Library Association 2008; 96(4): 335-342.

[14] Executive Order Number 13166. (Aug. 16, 2000). "Improving Access to Service for Persons with Limited English Proficiency." 65 Federal Register 50121.

[15] US Centers for Disease Control and Prevention (CDC). (1994). National Public Health Performance Standards Program (NPHPSP). National Public Health Performance Standards Program. Retrieved from http://www.cdc.gov/nphpsp/essentialServices.html

[16] National Cancer Institute (NCI). (2007). Health Information National Trends Survey. Retrieved from http://hints.cancer.gov/

[17] Semansky RM, Altschul D, Sommerfield D, Hough R, Willging CE. (2009). Capacity for delivering cultur-ally competent mental health services in New Mexico: Re-sults of a statewide agency survey. Administration and Policy in Mental Health and Mental Health Services Research 2009; 36(5): 289–307.

[18] Turner AM, Bergman M, Brownstein M, Cole K, Kirchhoff K. A Comparison of Human and Machine Translation of Health Promotion Materials for Public Health Practice: Time, Costs, and Quality. Journal of Public Health Management and Practice 2014; 20(5): 523-529.

[19] European Commission. (2001). European Multilingual Thesaurus on Health Promotion (Spanish). Retrieved from http://aims.fao.org/kos/european-multilingual-thesaurus-health-promotion

[20] Marchionini G. (2001). Planning new Internet services: Tracking public need, task, and expectation trajectories. Fi-nal Report to the Bureau of Labor Statistics. Retrieved from http://ils.unc.edu/~march/BLS_final_report_2001.pdf

[21] Trafford M, Consales J, Hamasu C. The role of information science and knowledge-based resources in de-livering telehealth services. Paper presented at Pacific Medical Technology Symposium, 1998. Honolulu, HI.

[22] Singh PM, Wight CA, Sercinoglu O, Wilson DC, Boytsov A, Raizada MN. Language preferences on websites and in Google searches for human health and food information. Journal of Medical Internet Research 2007; 9(2): e18.

[23] Berland GK, Elliott MN, Morales LS. Health information on the Internet: Accessibility, quality, and readability in English and Spanish. Journal of American Medical Association 2001; 285(20): 2612-2621.

[24] Matías-Carrelo LE, Chávez LM, Negrón G, Canino G, Aguilar-Gaxiola S, Hoppe S. The Spanish translation and cultural adaptation of five mental health out-come measures. Culture, Medicine, and Psychiatry 2003; 27(3): 291-313.

[25] Turner AM, Mandel H, Capurro D. Local Health Department Translation Processes: Potential of Machine Translation Technologies to Help Meet Needs. AMIA Annual Symposium Proceedings 2013; 1378-1385.

[26] Yakima Health District. (2012). Yakima Public Health. Retrieved December 5, 2012, from http://yakimahealthdistrict .org

[27] Kirchhoff K, Turner AM, Axelrod A, Saavedra F. (2011). Application of statistical machine translation to public health information: A feasibility study. Journal of American Medical Informatics Association 2011; 18(4): 473-478.

[28] Turner AM, Bergman M, Brownstein M, Cole K, Kirchhoff K. A comparison of human and machine translation for public health practice: time, costs and quality. J Pub Health Manag Pract 2014; 20(5): 523-9.

[29] Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminolo-gies: Report on the results of the NLM/AHCPR large-scale vocabulary test. Journal of American Medical Informatics Association 1997; 4(6): 484-500.

[30] Resnick MP, Ripple AM, Auston I, Rindflesch TC. An ontology for public health to support enhanced information retrieval. American Medical Informatics Association Annual Symposium Proceedings 2010, Washington, DC.

Address for correspondance

Daniel Capurro, MD, PhD, Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile. Lira 63, Santiago. dcapurro@med.puc.cl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-386

Serious Games: A Concise Overview on What They Are and Their Potential Applications to Healthcare

Guido Giunti[a], Analía Baum[a], Diego Giunta[a], Fernando Plazzotta[a], Sonia Benitez[a], Adrián Gómez, Daniel Luna[a], Fernan González Bernaldo de Quiros[a]

[a]*Hospital Italiano de Buenos Aires, Buenos Aires, Argentina*

Abstract

Younger generations are extensive users of digital devices; these technologies have always existed and have always been a part of their lives. Video games are a big part of their digital experience. User-centered design is an approach to designing systems informed by scientific knowledge of how people think, act, and coordinate to accomplish their goals. There is an emerging field of intervention research looking into using these techniques to produce video games that can be applied to healthcare. Games with the purpose of improving an individual's knowledge, skills, or attitudes in the "real" world are called "Serious Games". Before doctors and patients can consider using Serious Games as a useful solution for a health care-related problem, it is important that they first are aware of them, have a basic understanding of what they are, and what, if any, claims on their effectiveness exist. In order to bridge that gap, we have produced this concise overview to introduce physicians to the subject at hand.

Keywords:

Serious games, games for health, patient education, user-centered design, mhealth, behavioral health, medical education.

Introduction

Younger generations are extensive users of digital devices[1] to the extent of being designated by some as ''digital natives''[2,3]. For them, these technologies have always existed and have always been a part of their lives[4]. Video games are a big part of their digital experience. Already in 2004, the average video game player was aged 30 years old and had played computer games for almost 10 years[5]. The average child aged 8–10 years old spent 65 minutes per day playing video games; 52 minutes/day among youth aged 10–14 years and 33 minutes/day among teenagers aged 15–18 years[1]. Today, most of these stats have remained the same or increased in number[6]. Video games reach a large and diverse audience who expect extended contact; suggesting games can attract and maintain attention, a key component for effective behavior change[7].

User-centered design (UCD) is an approach to designing systems informed by scientific knowledge of how people think, act, and coordinate to accomplish their goals[8]. UCD design practices employ both formative and summative practices in order to achieve systematic discovery of useful functions grounded in an understanding of the work domain. The user experience of video games has itself become a substantial topic of human computer interaction, with researchers developing models and methods as well as heuristics for the usability or playability of games[9–11]. There is an emerging field of intervention research looking into using these techniques to produce video games that can be applied to healthcare. Games with the purpose of improving an individual's knowledge, skills, or attitudes in the "real" world are called "Serious Games"[12]. Serious Games applied to medical or health-related purposes are growing rapidly in number and in types of applications; however, physicians might not be aware of such developments.

Before doctors and patients can consider using Serious Games as a useful solution for a health care-related problem, it is important that they first: are aware of them, have a basic understanding of what they are, and what, if any, claims on their effectiveness exist. In order to bridge that gap we have produced this concise overview to introduce physicians to the subject at hand. We will first explain what Serious Games are and how they are fundamentally different from entertainment games. Secondly, we will provide a theoretical framework to understand why games and gaming can be appealing. Thirdly, we will dissect and elaborate on the elements and principles that make up Serious Games. Fourthly, we will summarize some examples and success cases of serious game design used to promote skill acquisition or modify behavior in other health and educational domains. Finally, we discuss and provide recommendations for future intervention research regarding the use of Serious Games to aid and promote health care.

Definitions

Humanity has played games since prior to written history[13], suggesting that playing games meets enduring psychological needs[14]. A game is a physical and mental contest with a goal or objective, played according to a framework, or rules, that determines what a player can and cannot do inside a game world[15].

A video game is any game played on a digital device and encompasses a wide range of games played at arcades, over the Internet on personal computers, or on dedicated game consoles (e.g., Nintendo Wii, Sony PlayStation, or Microsoft Xbox) or handheld units (e.g., Smartphones, Nintendo Game Boy, Sony PSP). To win the game, video games challenge players to use the information they obtain as they navigate the game world[16,17] thereby providing an interesting education and training modality[18].

The emerging genre of "serious video games" or "Serious Games" employs the medium's rich, role-playing, story-based

environments to teach, train, and change knowledge, attitudes, and behavior[19]. Serious Games are often referred to as games for health when they target health behaviors[20]. Unlike regular video games, Serious Games have the dual goal of entertaining, while promoting behavior change[12,21]. Achieving the proper balance between "fun-ness" (ie, components that entertain) and "serious-ness" (ie, the components that promote behavior change) is both essential and difficult to attain [12].

The Theory Behind Serious Games

Games are played primarily for entertainment or "fun,"[22] but what constitutes "fun" is not well understood. Serious game design draws from a large body of empirical research suggesting that learning is maximized when it is occurs in relevant contexts that engage learners[23]. They employ principles of video game design to create enjoyable and immersive environments but, importantly, they are also grounded in theories of learning and development[24]. Serious Games focus on providing feedback related to achieving long-term goals and enhancing intrinsic motivation for learning by providing players with information about their progress toward incremental and primary learning goals[25].

Reviewing the literature on the subject, we find that there are several theories and models used to explain the appeal of games. Next we will present a small selection of theoretical models that often show up in literature regarding Serious Games.

Self-determination theory (SDT) is a macro-theory of human motivation that has been applied to identifying which factors sustain individuals' motivation within video games[26]. SDT postulates that the more often basic psychological needs for autonomy, competence, and relatedness are satisfied within a game context, making both the experience more enjoyable and the motivation more sustainable[27]. When considering games as a set of behavior changes, Social Cognitive Theory (SCT) is another commonly cited theory[7]. SCT proposes that behavior change is a function of enhanced skills and confidence (self-efficacy) in doing the new behavior[7], while modeling[7] and feedback[28] are keystones for learning skills. A comprehensive model of learning for behavior change in video games is based on SCT and the elaboration likelihood model[29] and includes the following steps: attention, retention, production, and motivation.

The role of play in learning is informed by Vygotsky's social constructivist theory[30,31] and the concept of ''flow'' theory in task engagement[32], related to providing achievable challenges in learning, are often used. For example, flow theory suggests that engagement in learning is highest when perceived challenges and skills are well matched[32]. Garris et al.[19] suggest that learning is enhanced when participants discover and use information rather than memorize it.

In sum, serious game design merges learning theory and empirical findings about maximizing skill learning and generalization of learning together with principles of game design to create a unique intervention tool that can target any set of cognitive, social, affective, and/or health-related skills with the goal of improving outcomes beyond the context of the game.

Dissecting Serious Games: Principles and Elements

Hunicke et al.[33] developed a game design framework called MDA (signifying Mechanics, Dynamics, and Aesthetics), to help understand games. Fundamental to this framework is the idea that games are more like artifacts than media, they need to be thought of in terms of the behavior they produce via interaction.

Jesse Schell in his book "The Art of Game Design: a book of lenses"[34] and Ralph Koster in "A Theory of Fun for Game Design"[35] point out common video game components such as immersive storylines, characters, goals directed around targeted skills, rewards and feedback about goal progress, increasing levels of difficulty, and the provision of choice. These same elements are highlighted by Baranowski et al[20] and Kapp et al.[25] in their work on Serious Games.

Storylines engage individuals by means of their empathy with the protagonist and enable individuals to experience content in meaningful contexts[20]. In a serious game, the story narrative is built to support learning of the specific educational content targeted by the intervention[36]. This ''interactive storytelling'' technology allows for players' interactions with computer-controlled characters and subsequent decisions within the game to shape the goals and outcome of the storyline[37]. Characters in these stories can be protagonists, who serve as models, and antagonists, who attempt to impede the protagonists, thus adding tension and conflict that act as the motivating factor behind the story's action and plot[12,20]. Character modeling and dialogue can convey knowledge, demonstrate skills, and enhance self-efficacy[38].

Goals in games are objectives that players need to accomplish to succeed. These changes in behavior convey a statement of intention and give focus and direction to efforts[39]. Goals provide a standard, or benchmark, against which goal attainment can be assessed[40]. Continuous feedback and rewards for progress are critical for shaping behavior in Serious Games as learners work towards achieving challenging goals. In designing rewards and feedback, both intrinsic and extrinsic motivation needs to be considered[41].

As suggested earlier, SDT states that provision of choice is one of the important tools of fostering intrinsic motivation and enjoyment in Serious Games[14,42]. In addition to providing individualized levels of difficulty, provisions of choice within a serious game can allow learners to maintain a sense of autonomy and control over their learning experience[43].

Healthcare Applications

Serious Games have been shown in a systematic review to be at least as effective as conventional tests in improving cognitive abilities in the elderly[44]. In a randomized controlled trial (RCT) they seemed more effective than conventional neuropsychological interventions when it comes to improving neuropsychological abilities of alcoholic patients[45].

There are RCTs of computer-based interventions being used to improve emotion and face identity recognition abilities in autism[46] and meta-analyses on language and social skills[47] with the goal of improving psychosocial outcomes in both mental health and developmental disorders.

Serious game-based interventions have been used in RCTs to support rehabilitation in disabled patients, showing equal effectiveness compared to conventional training programs[48].

Games have been applied to promote healthy behavior in children[49] and to educate patients[50]. Serious game-based patient education has also been shown to increase the treatment adherence among adolescents with leukemia in an RCT [51]. The characters in educational games can include mentors who facilitate learning by providing guidance in the game[36,52].

Serious Games have been shown in RCTs to add to training medical personnel[53] and reviews show improvement of understanding of geriatric principles among medical students compared to conventional training methods[54]. One strategy for personalizing training/game play is to use adaptive progressions. In other words, on a trial-by-trial basis, the level of difficulty of the training is specifically adapted to the player's in–the-moment game performance.

Conclusions and Directions for Future Research

Video games are an experiential activity, rather than a presentation requiring memorization or assimilation of out-of-context facts. Video games can promote "situated learning"[55] in which players discover and learn through exploration and experimentation[56,57]. Through gameplay, players "vicariously" experience desirable and undesirable consequences without putting themselves in harm's way[57]. When using Serious Games in health care, end users (clinicians, patients, or educators) must decide whether games are safe and effective enough to be used for their intended purposes. In order to do so, they need consistent, transparent, and reliable assessments, however, studies on Serious Games'validity and effectiveness remain scarce[53,58].

Lewis[59] and Albrecht[60] have recently published guidelines reporting standards to support clinicians and patients in distinguishing high quality mHealth apps yet these have two important shortcomings when it comes to games. First, explicit information on a serious game's content and didactic features is required, as the external purpose of a serious game is frequently less obvious to the user than in the case of mHealth apps. Second, Serious Games require additional validation steps (eg, construct and predictive validity), compared to non-interactive information platforms. Gameplay is dynamic and learning goals in gameplay are often not disclosed to the user. In fact, the user learns by playing the game, whereas discovery in itself may be part of the gameplay. Disclosing learning goals could thus be counterproductive[61].

The Dutch Society for Simulation in Healthcare (DSSH) developed a consensus-based framework[62] based on Lewis'[59] and Albrecht's[60] work. This framework could be a valid tool to assess Serious Games but further exploration is required. It should also be considered that a game's validity does not predict a game's success nor its attractiveness to the user, which also depends on its entertainment capability and distribution method[63].

There is a need for more literature on Serious Games and their potential in informatics specific areas as well.

Finally, player expectations should also be considered. When faced with regular games players simply expect to be immersed in the story or gameplay of the product. Players may have different expectations for health video games and might be more critical as to the "fun" they are having.

Serious Games' potential make them something that physicians should be aware of but their complexity and lack of valid ways of assessment can play against their widespread use. Regardless, Serious Games are a promising tool which, if properly crafted, could be used to create high-impact interventions.

References

[1] Rideout VJ, Roberts DF, Foehr UG. Generation M: Media in the lives of 8-18 year olds [Internet]. Victoria. 2005 p. 85. Available from: http://www.kff.org/entmedia/upload/8010.pdf\nhttp://www.kff.org/entmedia/upload/Generation-M-Media-in-the-Lives-of-8-18-Year-olds-Report.pdf\nL:_Synch_folder\Reference Manager Data\Readings\rideout_roberts_foehr2005.pdf

[2] Prensky M. Digital Natives,Digital Immigrants Part 1. Horiz [Internet]. 2001;9:1–6. Available from: http://www.emeraldinsight.com/10.1108/10748120110424816

[3] Prensky M. Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently? On the Horizon. 2001. p. 1–6.

[4] Johnson L, Smith R, Willis H, Levine A, Haywood K. The 2011 Horizon Report [Internet]. Media. 2011 p. 36. Available from: http://wp.nmc.org/horizon2011/

[5] Entertainment Software Association. Essential facts about the computer and video game industry. 2005.

[6] Entertainment Software Association. Essential Facts about the Computer and Video Game Industry. 2014.

[7] Bandura A. Social foundations of thought and action : a social cognitive theory. Englewood Cliffs, N.J: Prentice-Hall, 1986. xiii, 617 pp. 1986.

[8] Flach JM, Dominguez CO. User-centered design: Integrating the User, Instrument, and Goal. Ergon Des. 1995;(July):19–24.

[9] Shaffer N. Heuristic Evaluation of Games. Game Usability: Advancing the Player Experience. 2008. p. 79–89.

[10] Bernhaupt R. Evaluating User Experience in Games: Concepts and Methods. HumanComputer Interaction Series. 2010.

[11] Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Comput Entertain. 2005;3:1–24.

[12] Thompson D, Baranowski T, Buday R. Serious Video Games for Health How Behavioral Science Guided the Development of a Serious Video Game. Simul Gaming [Internet]. 2010 Aug 1;41(4):587–606. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2919172&tool=pmcentrez&rendertype=abstract

[13] Juul J. Video games between real rules and fictional worlds. Video games between real rules and fictional worlds. 2005.

[14] Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: A self-determination theory approach. Motiv Emot. 2006;30:347–63.

[15] Gillin JL, Huizinga J. Homo Ludens: A Study of the Play-Element in Culture. American Sociological Review. 1951. p. 274.

[16] Lujan HL, DiCarlo SE. Too much teaching, not enough learning: what is the solution? Adv Physiol Educ. 2006;30:17–22.

[17] Barnett DJ, Everly GS, Parker CL, Links JM. Applying educational gaming to public health workforce emergency preparedness. American Journal of Preventive Medicine. 2005. p. 390–5.

[18] Vilozni D, Barker M, Jellouschek H, Heimann G, Blau H. An interactive computer-animated system (SpiroGame) facilitates spirometry in preschool children. Am J Respir Crit Care Med. 2002;164:2200–5.

[19] Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simulation & Gaming. 2002. p. 441–67.

[20] Tom Baranowski, PhD, Richard Buday, FAIA, Debbe I. Thompson, PhD and J. Playing for Real: Video Games and Stories for Health-Related Behavior Change. 2009;34(1):74–82.

[21] Thompson D, Baranowski T, Buday R, Baranowski J, Juliano M, Frazior M, et al. In pursuit of change: youth response to intensive goal setting embedded in a serious video game. J diabetes Sci Technol. 2007;1:907–17.

[22] Barendregt W, Bekker MM. Developing a coding scheme for detecting usability and fun problems in computer games for young children. Behav Res Methods. 2006;38:382–9.

[23] Catalano CE, Luccini AM, Mortara M. Best Practices for an Effective Design and Evaluation of Serious Games.

[24] De Freitas SI. Using games and simulations for supporting learning. Learning, Media and Technology. 2006. p. 343–58.

[25] Kapp KM. The Gamification of Learning and Instruction: Game- based Methods and Strategies for Training and Education. San Francisco, CA: Pfeiffer; 2012.

[26] Rigby S, Richard R. Immersion and Presence. Glued to Games: How Video Games Draw Us In and Hold Us Spellbound. 2011. p. 81–96.

[27] Sylvester BD, Standage M, Dowd a J, Martin LJ, Sweet SN, Beauchamp MR. Perceived variety, psychological needs satisfaction and exercise-related well-being. Psychol Health. 2014 Jan;29(9):1044–61.

[28] Kreuter MW, Farrell DW, Olevitch LR, Brennan LK. Tailoring health messages: customizing communication with computer technology. Lawrence Erlbaum Associates; 2000.

[29] Thompson D, Baranowski J, Cullen K, Baranowski T. Development of a theory-based internet program promoting maintenance of diet and physical activity change to 8-year-old African American girls. Comput Educ. 2007;48:446–59.

[30] Barab S, Thomas M, Dodge T, Carteaux R, Tuzun H. Making learning fun: Quest Atlantis, a game without guns. Educational Technology Research and Development. 2005. p. 86–107.

[31] Vygotsky LS. Mind in society: The development of higher psychological processes [Internet]. Mind in Society The Development of Higher Psychological Processes. 1978. Available from: http://www.amazon.com/dp/0674576292

[32] Shernoff D, Csikszentmihalyi M, Schneider B, Shernoff E. Student engagement in high school classrooms from the perspective of flow theory. Sch Psychol Q [Internet]. 2003;18:158–76. Available from: http://www.researchgate.net/publication/232520082_Student_engagement_in_high_school_classrooms_from_the_perspective_of_flow_theory

[33] Hunicke R, Leblanc M, Zubek R. MDA : A Formal Approach to Game Design and Game Research.

[34] Schell J. The Art of Game Design. Elsevier; 2008.

[35] Koster R. A theory of Fun for Game Design. Paraglyph Press; 2005.

[36] Lu AS, Thompson D, Baranowski J, Buday R, Baranowski T. Story Immersion in a Health Videogame for Childhood Obesity Prevention. Games Health J [Internet]. 2012;1:37–44. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3779587&tool=pmcentrez&rendertype=abstract

[37] Klimmt C, Roth C, Vermeulen I, Vorderer P, Roth FS. Forecasting the Experience of Future Entertainment Technology: “Interactive Storytelling” and Media Enjoyment. Games and Culture. 2012. p. 187–208.

[38] Schunk DH. Vicarious Influences on Self-Efficacy for Cognitive Skill Learning. J Soc Clin Psychol [Internet]. 1986 Sep [cited 2014 Dec 22];4(3):316–27. Available from: http://guilfordjournals.com/doi/abs/10.1521/jscp.1986.4.3.316

[39] O’Connor TJ, Cooper RA, Fitzgerald SG, Dvorznak MJ, Boninger ML, VanSickle DP, et al. Evaluation of a manual wheelchair interface to computer games. Neurorehabil Neural Repair. 2000;14:21–31.

[40] Fitzgerald SG, Cooper RA, Thorman T, Cooper R, Guo S, Boninger ML. The GAME(Cycle) exercise system: comparison with standard ergometry. J Spinal Cord Med. 2004;27:453–9.

[41] Habgood MPJ, Ainsworth SE. Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. J Learn Sci [Internet]. 2011 Apr 19;20(2):169–206. Available from: http://www.tandfonline.com/doi/abs/10.1080/10508406.2010.508029

[42] Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol. 2000;55:68–78.

[43] Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Review of General Psychology. 2010. p. 154–66.

[44] Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: A systematic review. PLoS One. 2012;7.

[45] Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, et al. Executive functioning in alcoholics following an mhealth cognitive stimulation program: Randomized controlled trial. J Med Internet Res. 2014;16.

[46] Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L, et al. Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! program. J Child Psychol Psychiatry Allied Discip. 2010;51:944–52.

[47] Grynszpan O, Weiss PLT, Perez-Diaz F, Gal E. Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. Autism [Internet]. 2014;18:346–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24092843

[48] Prange GB, Kottink AIR, Buurke JH, Eckhardt MMEM, van Keulen-Rouweler BJ, Ribbers GM, et al. The Effect of Arm Support Combined With Rehabilitation Games on Upper-Extremity Function in Subacute Stroke: A Randomized Controlled Trial. Neurorehabil Neural Repair [Internet]. 2014; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24878589

[49] Majumdar D, Koch P a, Lee H, Contento IR, Islas-Ramos ADL, Fu D. "Creature-101": A Serious Game to Promote Energy Balance-Related Behaviors Among Middle School Adolescents. Games Health J [Internet]. 2013;2:280–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24761326

[50] Cooper H, Cooper J, Milton B. Technology-based approaches to patient education for young people living with diabetes: A systematic literature review. Pediatric Diabetes. 2009. p. 474–83.

[51] Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics [Internet]. 2008 Aug [cited 2014 Dec 17];122(2):e305–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18676516

[52] Thompson D, Baranowski T, Buday R. Conceptual Model for the Design of a Serious Video Game Promoting Self-Management among Youth with Type 1 Diabetes. 2010;4(3):744–9.

[53] Graafland M, Schraagen JM, Schijven MP. Systematic review of Serious Games for medical education and surgical skills training. Br J Surg [Internet]. 2012 Oct [cited 2014 Jul 19];99(10):1322–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22961509

[54] Lagro J, van de Pol MHJ, Laan A, Huijbregts-Verheyden FJ, Fluit LCR, Olde Rikkert MGM. A Randomized Controlled Trial on Teaching Geriatric Medical Decision Making and Cost Consciousness With the Serious Game GeriatriX. J Am Med Dir Assoc [Internet]. Elsevier Ltd; 2014 Dec [cited 2014 Dec 21];15(12):957.e1–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24913210

[55] Anderson JR, Reder LM, Simon HA. Situated Learning and Education. Educ Res. 1996;25:5–11.

[56] Squire K. From Content to Context: Videogames as Designed Experience. Educational Researcher. 2006. p. 19–29.

[57] Shaffer DW, Squire KR, Halverson R, Gee JP. Video games and the future of learning. Phi delta kappan [Internet]. 2005;87:104–11. Available from: http://www.wcer.wisc.edu/publications/workingpapers/working_paper_no_2005_4.pdf

[58] Van Velsen L, Beaujean DJMA, van Gemert-Pijnen JEWC. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak [Internet]. 2013;13:23. Available from: http://www.biomedcentral.com/1472-6947/13/23

[59] Lewis TL. A systematic self-certification model for mobile medical apps. Journal of medical Internet research. 2013.

[60] Albrecht U-V. Transparency of health-apps for trust and decision making. J Med Internet Res [Internet]. 2013 Jan [cited 2014 Dec 22];15(12):e277. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3958693&tool=pmcentrez&rendertype=abstract

[61] Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to Systematically Assess Serious Games Applied to Health Care. JMIR Serious Games [Internet]. 2014 Nov 11 [cited 2014 Nov 12];2(2):e11. Available from: http://games.jmir.org/2014/2/e11/

[62] Dutch Society for Simulation in Healthcare [Internet]. Available from: http://www.dssh.nl/en/news/first-international-quality-labels-for-valid-games-granted

[63] Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of Serious Games: An overview. Advances in Human-Computer Interaction. 2013.

Address for correspondence

drguidogiunti@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-391

Living with Lung Cancer – Patients' Experiences as Input to eHealth Service Design

Maria Hägglund[a], Peter Bolin[a, b], Sabine Koch[a]

[a] *Health Informatics Centre, department of Learning Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden*
[b] *Stockholm County Council, Sweden*

Abstract

The objective of the study is to describe the lung cancer care process as experienced by patients, as well as to perform a qualitative analysis of problems they encounter throughout the patient journey. A user-centered design approach was used and data collected through two focus group meetings with patients. We present the results in the form of a patient journey model, descriptions of problems related to the journey as expressed by patients and proposed eHealth services discussed by patients in the focus groups. The results indicate that not only is the patient journey fragmented and different for each patient going through it depending upon their specific type of lung cancer and treatment options, but their experiences are also highly individual and dependent on their personal needs and interpretations of the process. Designing eHealth to improve the patient journey will therefore require flexibility and adaptability to the individual's needs.

Keywords:

Consumer health informatics; Patient journey mapping; eHealth; Participatory design.

Introduction

Consumer health informatics [1] is a growing field of research, as more and more applications are developed for patients and citizens rather than for health care professionals. The term eHealth was introduced by Eysenbach in the year 2000 as "an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies" [2]. eHealth has the potential to revolutionize the way health care and prevention is provided, shifting the balance of power and responsibility from healthcare professionals to patients and citizens [3, 4]. Yet many applications developed for patients are either designed from a healthcare provider's perspective, e.g. applications to collect patient reported outcomes, or stand-alone health applications, e.g. mobile apps for activity tracking. A more balanced way for initiating eHealth service design taking patients' experiences of the patient journey into account is suggested in this study. To design eHealth services that provide patients with a holistic overview of their often fragmented care requires a deep understanding of their experiences of the patient journey. In service design [5], customer journey mapping is often used to capture the consumers' experiences of using a service, and this method has lately also been applied in healthcare to describe the patients' experiences [6, 7].

My Care Pathways

The work presented in this study was performed within the Swedish research project "My Care Pathways" [8]. The project aims to create new mobile citizen e-services that allow patients to follow, own, and manage their care process related information. The project also aims to adapt and further develop the Swedish National platform for citizen e-services and provide an open software development kit (SDK) for developing new e-services [9].

In the initial stages of the project, three patient groups were involved in the e-service design: stroke patients [10], lung cancer patients and patients undergoing planned hip surgery. In this paper, we focus on the lung cancer patients. The objective of the study is to describe the process of being diagnosed with and treated for lung cancer as experienced by patients, as well as a qualitative analysis of problem areas in this care process. In this paper, we present the patient journey model, examples of problems as expressed by patients as well as examples of proposed eHealth services to address them.

Methods

We applied a user-centered design approach [11] to the analysis of problems and needs and actively involving patients in the process. An overview of the approach is shown in Figure 1.

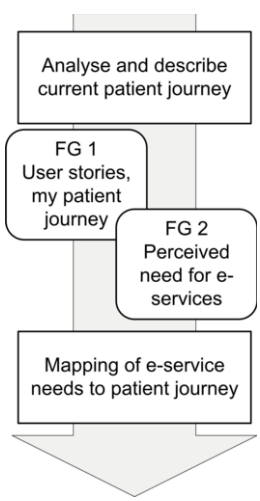


Figure 1 – Overview of the Needs Analysis Process

The first stage of the process was to do an initial analysis and description of the current patient journey. The initial patient journey model was based on literature and materials gathered from clinicians at the Karolinska hospital who are involved in different stages of the lung cancer diagnosis and treatment process. The model was then presented to the patients to validate that we had indeed captured the stages that are important for the patients.

Two focus group interviews were held to explore the problems and experiences of the patients and how these problems change throughout the patient journey. An overview of participants is given in Table 1. All participants were recruited by convenience sampling via the lung cancer patient organization Stödet (http://stödet.se) in Stockholm, Sweden.

Table 1- An overview of participants

Focus Group	Number of participants
FG 1	Patients (n = 4)
FG 2	Patients (n = 5)

All focus group meetings were facilitated by a moderator (first author of this paper, MH) and notes were taken by two researchers (PB, SK). Each focus group lasted 2-3 hours. The focus group meetings were audio recorded and transcribed. Content analysis [12] was used to identify categories and themes related to the patient experience. In addition, current paper-based information given to patients at different stages was gathered and studied.

The collected qualitative data was used to model a patient journey, referring to "the experiences and processes the patient goes through during the course of a disease and its treatment" [7]. The patient journey model aims to provide a common picture of the processes and the way the patients experience them.

Ethical approval for the study was obtained from the regional ethical review board (2011/2093–31/5).

Results

The results are presented in the form of a patient journey model, important patient experiences and proposed eHealth services.

The Patient Journey Model

We distinguish between phases and events in the patient journey model (Figure 2). A phase is extended over time and may incorporate several events. An event is a specific interaction between patient and healthcare, where information is created, shared or communicated. Patient journey models often only include phases, but to use the patient journey as a basis for design of eHealth services, it is important to also map these events, typically called touch points in customer journey maps [5].

We identified 5 distinct phases that the lung cancer patients go through: (1) pre-diagnosis care (primary and/or acute), (2) diagnostic examinations, (3) treatment, and finally (4) rehabilitation (when the patient is in remission), or (5) palliative care.

The *pre-diagnosis care (primary and/or acute)* phase can be a long and often uncertain process. Patients sometimes seek care on several occasions before the suspicion of lung cancer is actually raised, and a high proportion of patients who are finally diagnosed present with advanced disease [13, 14]. The first contact is often with a primary care physician, and if the suspicion of lung cancer is raised in primary care, referrals are sent for further assessments. However, it is not uncommon that patients seek acute care after having been misdiagnosed and treated with antibiotics in primary care and the referral to specialist care is sent from the acute care department [14]. An important event that all participants in this study described was when the decision was made to refer to diagnostic examinations.

The second phase, *diagnostic examinations and specialist assessment*, can be fragmented since many different clinics are involved in the examinations, which may include laboratory tests, chest X-ray, computer tomography (CT scan, CAT scan), positron emission tomography (PET scan), sputum cytology, fine-needle aspiration biopsy of the lung, bronchoscopy and many more. This phase is often perceived as time consuming and frustrating by the patients, as they are anxiously awaiting the results of the examinations. The phase is often coordinated from a pulmonary clinic, and begins when the patient is referred there for diagnosis and staging. After lung cancer has been diagnosed, the phase continues with further examinations to determine the type of lung cancer (non-small cell and small cell lung cancer) and staging, i.e. to determine how far the

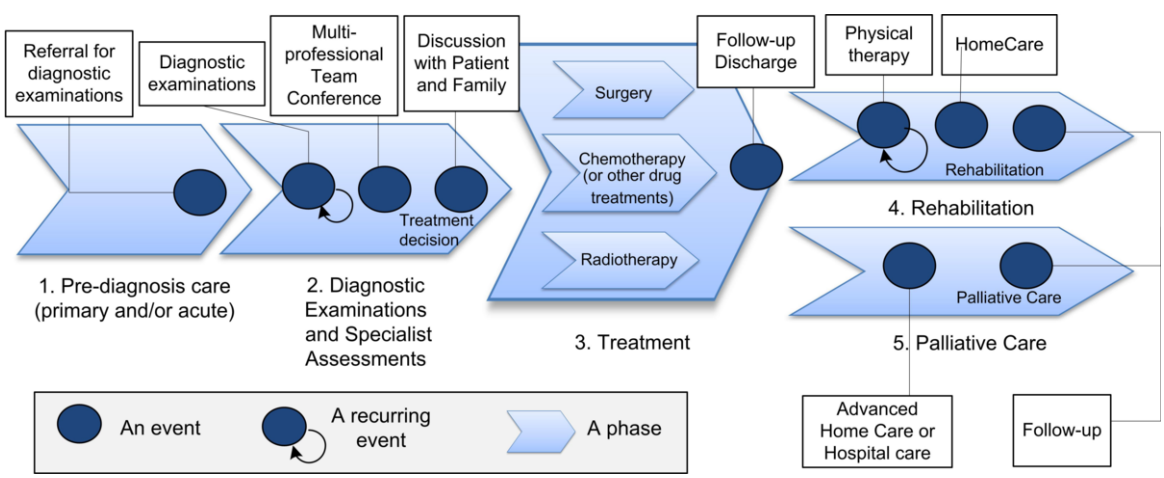


Figure 2 – The Lung Cancer Patient Journey Model

cancer has spread. The phase ends with a multi-professional conference where clinicians with different specialities meet to discuss the diagnosis and treatment options. The preferred option is then discussed with the patient and family members before treatment begins.

The *treatment* phase depends very much on the choice of treatment, but usually consists of surgery, radiotherapy, chemotherapy, or a combination of these. Different specialist clinics are responsible for different treatments, thorax for surgery, oncology for radiotherapy and chemotherapy, and pulmonary clinic for palliative and post-surgery care. Patient experiences varied depending on the type of treatment they had received. Patients who received care at different clinics expressed frustration at the poor communication between clinics and lack of coordination of activities. One participant who fell ill with breast cancer while undergoing treatment for her lung cancer described the lack of coordination:

"But I think the communication between the Radiology clinic and Thorax is poor. It feels like I have to… this spring when I was receiving treatment that would last until summer, I asked if it could clash with the medication I was getting from the Radiology clinic. […] The chemotherapy is a drug given for breast cancer too so they could 'take each other out' and it turned out no one told me to stop taking this medicine I was taking everyday. And they clashed. I got pretty ill."

Yet, once the treatment finally began, many patients felt relief, *"… it had been 3-4 months, I think, when they put the needle in my arm [for chemotherapy], and then I felt; now I will get better, now I have received help!"*

Not all patients undergo curative treatments, many lung cancer patients are diagnosed at a late stage and suffer from co-morbidities such as COPD (chronic obstructive pulmonary disease) or heart disease, so that curative treatments are deemed unwise and it is decided to go directly to palliative care instead. The treatments described above may still be used, to relieve symptoms, but not with a curative purpose. If treatments have been performed with a curative aim, the patient hopefully goes into remission and moves on to the rehabilitation phase, but if the treatments are not effective, a *palliative* phase begins with focus on pain management and quality of life.

Rehabilitation was an important topic for our participants, as it was not available to all patients automatically. Depending on which clinic the patient was treated at (oncology, thorax or pulmonary) rehabilitation was more or less available despite patients suffering from similar symptoms. This was expressed as frustrating and a sign of inequity.

The Patients' Experiences

Describing the process in terms of phases and events is important, but designers of eHealth services also need guidance when patients experience problems. Table 2 shows problems experienced by lung cancer patients related to the phases.

Table 2 – Patient Experienced Problems

Phase	Problem	Description
1	Unnecessary delays before diagnosis	Several of the participants expressed that it took too long for primary care to acknowledge their condition. One participant had to visit the emergency department and was then recommended to contact the pulmonary clinic directly. Participants felt that they had to fight to get to get the right care and that it depended on the individual healthcare professional they met.
2-3	Poor communication with health care	Participants expressed frustration at reaching the clinics at the hospital. Especially when receiving care at different clinics, finding contact information and reaching the right person was challenging. Different clinics at the same hospital also had different approaches to communicating with patients making it more difficult to manage.
2-3	Poor coordination	As both diagnostic and treatment activities are distributed on many different clinics, patients suffered from poor coordination when e.g. appointments were scheduled two days in a row rather than after each other on the same day. The patients also lacked information about who (which department/clinic) is responsible for which part of their treatment.
2-3	Understanding of procedures	Some of the procedures are major and the patients are naturally worried and concerned beforehand. An issue that was brought up in the focus groups was that patient information was often not up to date, or was provided in what they interpreted as "homemade" brochures. This did not improve the patients trust and confidence in health care.
4-5	Patient follow-up	The patients receive different types of patient surveys. They are often repetitive and far too long. As a patient you do not have the strength to answer the surveys all the time.
All	Poor support for multi-morbidity	Many lung cancer patients also suffer from other conditions, yet health care is not adapted to this. Frustration was expressed at having to repeat information and act as the coordinator between different clinics and clinicians.
All	Learning to manage one's care	Several participants expressed that it took time for them to understand where and how to find information and to manage their care. If the patient participated in a study, then they received more information, more continuity and more visits.
All	Lack of channels for giving feedback	At the same time, the patients lacked a channel for giving their feedback, either immediately after a meeting or long-term. Instead, the patient organization receives this information, without being able to pass it forward.
All	Understanding of rights to choose	The patients lacked information about what opportunities they have to e.g. choose doctor and treatments, second opinion. Trusting your physician was expressed as crucial, and several participants had asked to see another physician after a bad experience. However, this option was not something everyone was aware of, and it took some time for the participants to realize how important this was.

Examples of Proposed eHealth Services

Based on the modeled process and the identified problems, we suggested a number of potential eHealth services, and some examples of these are presented in Table 3.

One participant expressed an overall need for eHealth services in the following way:

"… I can't demand to have total control over the entire process, but I should at least have as much control so that I can trust health care and focus my energy on getting well".

It was important for the participants to have information and insight into the care processes, but they stressed that the responsibility needed to remain with health care – eHealth services should not be a means for health care to leave the responsibility for coordination and communciation to the patients.

Table 3 – Proposed eHealth Services

Problem	Proposed eHealth service
Overview of the diagnostic procedures	An e-service showing the **progress of the different diagnostic examinations** made was requested. The participants did not want to receive the results online, but wanted an expected time frame and did not want to have to wait for an appointment once the results were ready. *"So, it had taken time, but then I was supposed to wait 3 weeks for a doctor's appointment so I could find out what it was when they were done. Then I lost it completely..."*
Poor coordination	Participants suggested that health care should be working more with logistics to get the processes right. An **overview of the "normal" care process** would be useful to both patients and professionals. If it could also show the patient's current position and who is responsible for different parts of the process it would be even more useful.
Understanding the procedures	Available information beforehand is necessary to decrease anxiety. The information should be updated and easy to understand to make patients feel confident. Online information sites in various designs were recommended.
Patient follow-up	**Online patient surveys** which could make the answers available during the whole care process were requested.
Lack of channels for giving feedback	**e-services for giving immediate feedback** to healthcare were deemed useful by the participants, but they also wanted a way to communicate feedback later on when you have more perspective on the care process. Channels for making formal complaints are available, but the participants lacked an easy **means for communicating feedback and improvement suggestions**.
Poor communication with health care	**Improved services to communicate with health care** were requested. Individual patients have different preferences for synchronous and asynchronous communication, and alternatives should be available. However, overview and responsibility was again stressed – who do I contact about this specific issue?
Learning to manage one's care	Many participants expressed that it took them time to learn how to navigate through health care, where to find information and how to ensure you get the care you need. If this process could be faster it would be helpful, and support from other patients in combination with other **guiding or e-learning eHealth services** could be useful.
Understanding of rights to choose	e-services **clarifying patients' rights** both in terms of choosing a health care professional you trust, and when it comes to requesting second opinions, choosing treatment etc. Most participants expressed that they trusted the judgments made by health care professionals, especially during the multi-professional conference – but at the same time they had all experiences of questioning individual professionals' judgment and requesting to meet new physicians in the future.

The proposed eHealth services in Table 3 are not an exclusive list, but rather suggestions made by participants in the focus groups. Anyone reading about the problems may come up with their own ideas for eHealth services that could improve the situations for people living with lung cancer. This is one of the strengths of using patient journey mapping and patient experiences as a basis for design.

Discussion

We used the patient journey model to understand the processes a patient goes through before, during and after lung cancer treatment and the problems they experience during this journey. Gaining this insight is important input to proposing new eHealth services for lung cancer patients.

A limitation of this study is the low number of participants involved in the focus groups. Still, even this limited case study points to important insights into the patients' experiences, and important ideas for eHealth services were produced. These results will be used together with an in-depth content analysis of the patients' information and communication needs to further understand the content and functionalities eHealth needs to provide to support lung cancer patients throughout their patient journey.

All participants were recruited locally, and had received their care within the same county council. This limits the transferability of the results in terms of the patients' experiences and some of the proposed eHealth services. Yet, as a designer of eHealth services one may recognize similar problems from other contexts where these results can be applied. The important, transferable, result of this study is however the method of using patients' experiences as a basis for proposing new eHealth services to improve health care from a patient perspective. The approach can be applied locally in any context to explore patients' experiences and propose solutions to identified problems.

The lung cancer patient journey in Stockholm, Sweden, is fragmented and different for each patient going through it depending on their specific type of lung cancer and treatment options. In addition, their experiences are also highly individual and dependent on their personal needs and interpretations of the process. Designing eHealth to improve the patient journey therefore requires flexibility and adaptability to the individual's needs.

It is also important to acknowledge that health care is a socio-technical system [15, 16], and not all the problems identified in this study can be solved through new eHealth services. Therefore, it is crucial to involve healthcare professionals and other stakeholders within the healthcare organization to

address the problems from more organizational and work process perspectives too. In addition, when designing and implementing new eHealth services to support patients throughout the patient journey, it is imperative to consider what impact these will have of the surrounding organization and how do thenew tools for communication between patients and professionals affect the daily work.

Conclusion

Patient journey models and qualitative analysis of patients' experiences are powerful tools that can be used to improve health care from a patient perspective. In this study we show how such tools can be used as input to the design of eHealth services, but by creating a patient journey model and describing patients' experiences of going through this journey, we also create opportunities for reaching a common understanding of issues and problems experienced by patients, thereby facilitating improvement work and in the long run increased patient satisfaction.

The results indicate that not only is the patient journey fragmented and different for each patient going through it depending on their specific type of lung cancer and treatment options, but their experiences are also highly individual and dependent on their personal needs and interpretations of the process. Designing eHealth to improve the patient journey will therefore require flexibility and adaptability to the individual's needs.

Acknowledgments

The study is supported by VINNOVA – Swedish Agency for Innovation Systems through the project "My Care Pathways" (2011-02536). We also wish to thank all the participants in the study who shared their experiences openly with us.

References

[1] Eysenbach G. Consumer health informatics. BMJ Br. Med. J. 2000, vol. 320, pp. 1713–1716.

[2] Eysenbach G. What is e-health? J. Med. Internet Res. 2001 vol. 3, no. 2.

[3] Koch S, Hasvold T, Kushniruk AW, Marcelo A, and Kouematchoua Tchuitecheu G, eHealth - nurturing patient empowerment? State of the art and reflections from four continents. In: Panel discussion. Medinfo 2010, Cape Town, South Africa.

[4] Koch S, Improving quality of life through eHealth-the patient perspective. Stud. Health Technol Inform. 2012 vol. 180, pp. 25–29.

[5] Stickdorn M and Schneider J, This is service design thinking. John Wiley & Sons, Ltd, 2011.

[6] Trebble T, Hansi N, Hydes T, Smith M, and Baker M, Process mapping the patient journey: an introduction. BMJ 2010;341:c4078.

[7] Manchaiah V, Stephens D, and Meredith R, The patient journey of adults with hearing impairment: the patients' views. Clin. Otolaryngol., 2011 Jun;36(3):227-34

[8] Lundberg N, Koch S, Hägglund M, Bolin P, Davoody N, Eltes J, Jarlman O, Perlich A, Vimarlund V, and Winsnes C, My Care Pathways - creating open innovation in healthcare. Stud Health Technol Inform. 2013;192:687-91.

[9] "Health Innovation Platform (HIP)." [Online]. Available: http://www.healthinnovationplatform.se/en/. [Accessed: 10-Nov-2014].

[10] Hägglund M, Davoody N, Bolin P, Winsnes C, Lundberg N, Vimarlund V, and Koch S, Stroke Patients' Needs for Access to Care Process Related Information, in Medicine 2.0, 2012. http://www.medicine20congress.com/ocs/index.php/med/med2012/paper/view/1083

[11] ISO 13407, "Human-Centred Design Process for Interactive Systems." Geneva, 1999.

[12] Graneheim U H and Lundman B, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, Nurse Educ Today. 2004 Feb;24(2):105-12.

[13] Coleman M P, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan C E, Turner D, Marrett L, Gjerstorff M L, Johannesen T B, Adolfsson J, Lambe M, Lawrence G, Meechan D, Morris E J, Middleton R, Steward J, and Richards M, Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet. 2011 Jan 8;377(9760):127-38.

[14] Wilcock A, Crosby V, Freer S, Freemantle A, Caswell G, and Seymour J. Lung cancer diagnosed following emergency admission: a mixed methods study protocol to improve understanding of patients' characteristics, needs, experiences and outcomes. BMC Palliat Care. 2013 May 28;12:24.

[15] Baxter G and Sommerville I, Socio-technical systems: From design methods to systems engineering, Interacting with Computers, 2011;23:1, pp. 4–17

[16] Coiera E, Four rules for the reinvention of health care. BMJ. 2004 May 15;328(7449):1197-9.

Address for correspondence

Maria Hägglund, maria.hagglund@ki.se

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-396

Applying a Geospatial Visualization Based on USSD Messages to Real Time Identification of Epidemiological Risk Areas in Developing Countries: A Case of Study of Paraguay

Silvia Ochoa[a], Julia Talavera[a], Julio Paciello[a]

[a] *Polytechnic School, National University of Asuncion*

Abstract

The identification of epidemiological risk areas is one of the major problems in public health. Information management strategies are needed to facilitate prevention and control of disease in the affected areas. This paper presents a model to optimize geographical data collection of suspected or confirmed disease occurrences using the Unstructured Supplementary Service Data (USSD) mobile technology, considering its wide adoption even in developing countries such as Paraguay. A Geographic Information System (GIS) is proposed for visualizing potential epidemiological risk areas in real time, that aims to support decision making and to implement prevention or contingency programs for public health.

Keywords:
eHealth, mHealth, GSM, USSD, GIS.

Introduction

Health is one of the most critical areas regarding the use of information due to many existing regulations and primarily, public management.

In Latin America and the Caribbean, considerable inequity exists in the access to public health services . This is a result of a number of factors , including: lack of human resources, infrastructure, equipment, and medication. Further, a physical and cultural gap exists between the public health system and the population, related to low incomes and different ethnic backgrounds in developing countries. Thus, vulnerability is determined by income levels, geographic location, and ethnicity, causing an exclusion of millions of households in the region [1].

Information and Communication Technologies (ICT) increase the chance of improving the life quality in the community, as shown in previous works [2,3]. This paper focuses on applying ICT as a tool for public health surveillance. Thus, this initiative can be considered within the topic of *eHealth*. According to the *World Health Organization* (WHO) [4] eHealth is *the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including healthcare services, health surveillance, health literature, and health education, knowledge, and research.*

The purpose of the present work is to optimize the geolocated data collection process for suspected or confirmed disease occurrences through the use of mobile devices to enable the identification of potential epidemiological risk areas in real time. Visualization of collected data is proposed through a *Geographic Information System* (GIS).

The proposed model can be implemented via mHealth as an application that allows users to report disease occurrence from their mobile devices through the exchange of *Unstructured Supplementary Service Data* (USSD) messages. USSD is a feature included in *Global System for Mobile Communications* (GSM) mobile technology (known as "2G"), that allows bidirectional transmission of information between the mobile device and an operator-defined application. With the collected information, this work proposes a geospatial visualization to identify epidemiological risk areas. Also, it can be used as a supporting tool for improving the actions of the organizations responsible for health.

We propose involvement of the community members in data collection, using the community as an actor in the process of epidemiological surveillance. The model of *Community Health Workers* [5] arises as part of the community auxiliary health team, where workers are selected and trained in order to work for their own communities.

In the next sections of this paper, a description of the problem is presented, including statistical information about the penetration of technologies in Paraguay. Also, the proposed model and considered scenarios are presented in detail. Finally, a comparative analysis between the proposed model and other alternatives is presented, along with conclusions of the work.

Background

In Paraguay, the *General Office of Health Surveillance* (DGVS, for its acronym in spanish) of the *Ministry of Health and Welfare* (MSPBS, for its acronym in spanish) is responsible for providing continuous information about monitored diseases in order to support decision making regarding prevention, control, and elimination of these diseases.

All the information related to mandatory surveillance of diseases in Paraguay is centralized by DGVS. Every week, the *Data Management Unit* (UGD, for its acronym in spanish) receives a notification form – which is printed (fax or mail) or in digital format (e-mail) – from the health regions. The forms are received mainly in printed format. Each health region has an Epidemiological Unit, which is responsible for collecting information from all reporting units.

The UGD unit performs a manual process of digitizing and unifying the collected information in order to obtain an electronic spreadsheet, which indicates for each record the

reported disease and the specific health region during a epidemiological week, as defined by the MSPBS.

Finally, the information is sent to the *National Center of Communications* (CNE, for its acronym in spanish), where statistical analyses of the collected data are performed. An epidemiological newsletter is published weekly, showing the current status of the country. Typically, risk maps are used for visualizing the information.

According to statistical information from Paraguay in 2013 – with an estimated population of 6.700.000 – 95.5% owned a mobile device (97.1% urban areas, 93.4% rural areas). Additionally, 19.4% had landline communication subscription [6].

The General Office of Statistics and Census (DGEEC, for its acronym in spanish) announced that of all Paraguayan households, 26.6% have Internet-connected computers. However, it should be noted that Paraguayans have recently increased their access to the web using mobile devices (such as smartphones and tablets). For this reason it is estimated that the actual Internet penetration is approximately 30% [7].

From the mentioned statistics, the mobile technology – particularly GSM networks – represents a communication channel with a greater penetration in households than landlines or the Internet. This makes it an ideal tool to optimize the process of gathering information.

Based on these statistical values, we propose that an efficient way to penetrate more extensively in the territory of Paraguay is through GSM technology. Currently, internet and smartphone penetration remains insufficient in a developing country such as Paraguay, where the territory is mostly rural. Also, it should be considered that individuals typically excluded from public health services are those with low incomes or of different ethnicities.

Materials and Methods

Natural disasters, such as wildfires, floods, epidemics, and oil spills, can be modeled and presented through a GIS [8]. In public health, GISs have been used mainly for processing medical and epidemiological research that explores the magnitude and distribution of several health problems. In addition,, for analyzing, monitoring, and decision making related to this field.

The use of USSD communication protocol is considered, due to its ability to send information and instructions bidirectionally (between the mobile device and the applications that process the data). Thus, real time information can be provided by users, which will feed into a GIS, to allow the surveillance of epidemic risk areas. This will allow saving both economic and human effort, considering that, having this tool, the citizens and health workers would spend less time in the early identification of epidemiologic risk areas in order to support establishing policies, strengthen planning processes, and guiding actions to improve citizens' quality of life.

As shown in Figure 1, the information was collected in a centralized system, which may be accessible in real time to the actors involved in the information collection process. This will support the decision making at all levels and ensure that plans and policies are fully justified based on the evidence gathered. Thus, an improvement over the conventional non-automatic model was obtained, where the information originally reaches the levels required for decision making with a considerable delay.

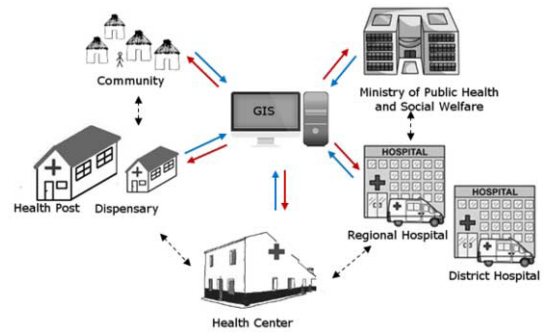


Figure 1 - Real time access to information

The posed problem in this work has been detached into two parts: data collection and visualization. A USSD module was designed for the first one, which provides greater coverage in terms of the number of citizens who will be able to access the service. For the second one, a GIS module was implemented, responsible for rendering (in real time) the information gathered for geographic visualization. The proposed model including the modules mentioned is shown in Figure 2.

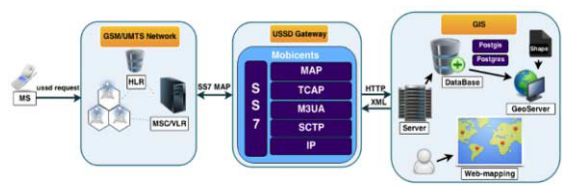


Figure 2 - Proposed model

The proposed model begins with the user or subscriber, who initiates a USSD dialogue by dialing a *Service Code* (SC) from the *Mobile Station* (MS). This message is encapsulated in a message *Mobile Application Part* (MAP) and is sent through the GSM network elements to the USSD Gateway (USSD module) which is responsible for receiving the request and identifying (based on the SC) which application has to be accessed in order to forward the request to it establishing a session. The application returns the initial menu to the USSD Gateway and this back to the user.

In this way, a USSD dialogue starts between the mobile device and the USSD application, where each entered option is logged in a georeferenced data base.

Finally, the data collected through the USSD module was processed and analyzed in the GIS module, which allows for visualizing real time information obtained from such data. In the following sections, both modules are explained in detail.

USSD Module

This module is responsible for data collection through a USSD application that uses RESTful web services to get an interactive menu of disease reports and registration.

The collected information was categorized according to the user who has made the notification. Every type of user will have a different notification accuracy level. This distinction has been deemed important because it is necessary to ensure the quality of the information. Inconsistent data or false positive notifications would not reflect reality, so the actions

taken or applied on these data cannot be used to improve or eliminate the identified health problems. Thus, the notifications can be visualized and compared in real time and for each user type. Among the possible types of users, the proposal includes: a citizen, a community health worker, and/or an official of the Ministry of Public Health. The last two types of users are able to implement more rigorous and reliable checks and verification to reduce false positives.

The Mobicents [9] platform was used to simulate the USSD dialogue, which is an implementation of *Open Source Software* (OSS) for development of USSD applications based on the *Signalling System No. 7* (SS7) stack protocol. In Figure 3, the message flow of a typical MAP protocol is presented for transferring data between a network node, the Mobicents platform, and the USSD application.

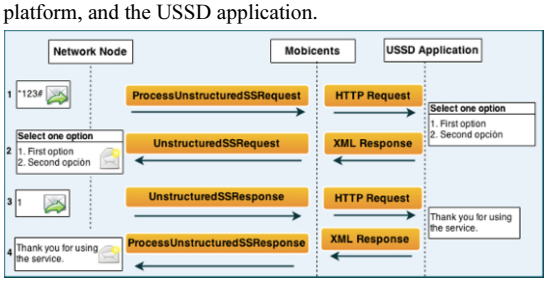


Figure 3 – USSD Dialog

The user dials the SC USSD code (e.g., *123#). A network node can represent the MSC, HLR, or VLR. The MAP protocol is used by the Mobicents platform to communicate with the network node and HTTP and XML protocols are used to communicate with the USSD application. The USSD application handles the business logic and sends the response back to the user in the same session.

GIS Module

The GIS Module is responsible for processing, analyzing, and categorizing data collected in the previous module, and allows the definition of thresholds on when to consider an area as being at risk for an epidemic. Finally, analysis of information is performed by a map viewer in real time.

This module aims to provide to the government, non-government organizations, and agencies that work on health services a tool to obtain georeferenced information of areas or regions affected by some kind of disease.

As shown in Figure 2, the GIS module consists of several components: a server application that is used by the USSD module to obtain interactive menus and persist the notifications of the users in the database, and a map server, which is responsible for analyzing the collected data and publishing the diseases layers with styles specific to each one of them. Finally, a web map visualization is included, where the available layers are passed to the map server and rendered to the final users.

Considered Scenarios

The scenarios considered in this work are differentiated by the locating mechanism used (that is, the mechanism that allows the mobile device's position to be determined). This information may be provided by the GSM network operator or by the user.

Automatic localization based on Cell-ID

There are several methods to determine the location of the MS. They can be different according to the technologies used to determine the location and the expected precision. In our scenario, the location method using the MS's Cell ID was used, considering its low cost and its inclusion in the GSM network, in order to determine the mobile device's location. The antenna identifier (the Cell ID) to which the device is connected is used to geographically locate the antenna, which is an approximation of the user's location.

For this scenario, the operator must provide geographic information of their *Base Transceiver Station* (BTS) antennas, so they can be identified by the visualization component of the GIS module. Since the information regarding the connected mobile device's antenna is handled by the telecommunication company, once a user's request is received, the USSD gateway simulates the consult to the HLR in order to determine the antenna to which the device is connected [10], and returns a MAP message with the Cell ID for that particular antenna.

The accuracy of this method depends on the cell size. We assumed 500 meters of accuracy within a city and about 10 kilometers in rural areas, representing an acceptable range considering the necessary resolution for identification of risk areas.

Manual localization

For this investigation two additional menu fields were included in the USSD application: the department (state) and city must be selected by the users in order to map to their corresponding latitude and longitude in the GIS, based on its own geographic data.

This scenario was considered as an alternative way to determine location when the system fails to obtain the proper information from the the telecommunication company. It is noteworthy that in these cases the accuracy of the model decreased significantly, and also required the user to know and select their department and city.

Results

The software evaluation process allows us to justify investing in the technology, based on the goals and needs of the consumer.

The development and selection of high quality software represents a critical point in the process, considering that the development and implementation of our method drives the success or failure of processes that are supported by these tools. This can be achieved by defining appropriate quality characteristics, taking into account the purpose and use of the software [11].

To verify the benefits and drawbacks of the conventional model currently used by the DGVS, and of the proposed model, a comparative study was conducted , considering each of the characteristics and subcharacteristics defined in Table 1.

The comparison is based on a set of general characteristics and subcharacteristics which were defined in order to fix the problems that emerged during the present study and considering the characteristics and subcharacteristics of ISO 25010 that are comparable to the evaluation models.

Table 1– Quality characteristics

Characteristic	Subcharacteristic	Evaluation	Weight	Conventional Model	Proposed Model
Territorial Scope	Data Collection	Complies with this subcharacteristic if data collection covers at least 70% of the population as potential notifiers.	1	0	1
	Disemination Information	Complies with this subcharacteristic if the model allows access to information to support decision making.	1	1	1
Availability of Information	Data Collection	Complies with this subcharacteristic if data collection is done in real time.	1	0	1
	Disemination Information	Complies with this subcharacteristic if the information is disseminated in real time.	1	0	1
Product quality	Functional Pertinence	This subcharacteristic counts one point for each tool supported by the model.	3	2	3
	Documentation	Complies with this subcharacteristic if the model has adequate documentation to facilitate the use of the product to the end user.	1	0	1
	Accessibility	Complies with this subcharacteristic if the product can be used by illiterate or blind users.	1	0	0
Reliability	Objectivity	Complies with this subcharacteristic when the information provided is not biased towards a particular opinion.	1	1	1
	Topicality	Complies with this subcharacteristic if the displayed information is updated on a weekly basis, considering the epidemiological calendar.	1	1	1
	Credibility	Complies with this subcharacteristic if is possible to determine that the information provided is reliable and relevant with respect to the subject matter.	1	1	0
Interoperability	Exchange Formats	This subcharacteristic counts one point for each format (GIS, Web Services, or images) supported by the model.	3	1	3
	Open Data	Complies with this subcharacteristic if the data are made openly available.	1	0	1

Discussion

The score obtained by the models evaluated in each characteristic can be seen in Figure 4, where the scores for our proposed model far exceed those of the conventional model. With regard to *Territorial Scope* – and considering the process of data collection through the use of USSD technology – the proposed model outperforms the conventional model. The *Availability of the Information* (in real time) is a characteristic of vital importance, because our study is within the health sector, and whose subcharacteristics are dependendientes. In other words, since the information is available in real time, data should be collected in real time. Considering *Product Quality* (specifically when users interact with the collection or visualization tools) the conventional model provides the option to view historical information from the epidemiological newsletter from previous weeks, but has no tools to interact with the information provided. Of the three subcharacteristics evaluated with respect to functional pertinence, the conventional model complies with two, while the proposed model complies with three. Additionally, the conventional model provides little documentation, which hinders the understanding and interpretation of risk maps displayed in the epidemiological newsletter. This is opposite for the proposed model, which displays official documents in the same application. Finally, considering accessibility, both the proposed model and the conventional model cannot be used by people with certain physical impairments, such as people who cannot read and/or write, or who are blind. A drawback of using the proposed model is its low score in *Reliability*. This is because the reporting user base is much broader than that of the conventional model. The *Interoperability* characteristic was not highly rated for the conventional model since it only allows the display of static information, which does not allow share and reuse of the information. Instead, the proposed model complies with all considered formats of information exchange, and also openly shares data.

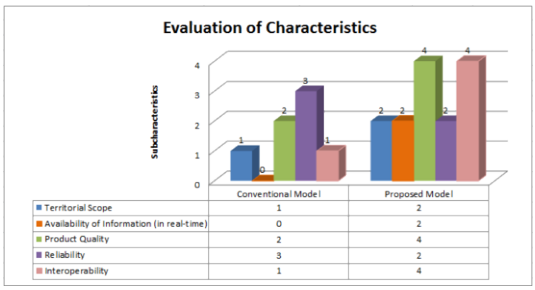


Figure 4 – Final Evaluation of Characteristics

Figure 5 shows the final scores of the models evaluated. The proposed model has a final score of 89.33%, the highest score, wheras the conventional model has a final score of 43%.

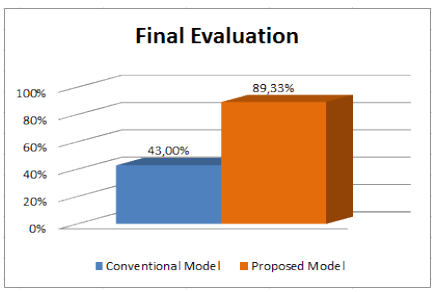


Figure 5 – Final Evaluation

Conclusion

This work highlighted the potential and benefits of a proposed model used for ICT-based epidemiological surveillance, which can easily be extended to other areas, towards the implementation of *Electronic Government* (eGov) and *Mobile Health* (mHealth) paradigms.

The emergence of new strategies for public health play an important role in Paraguay, for parties including primary health care (and family health units), community workers, and community empowerment towards personal health.

We have designed a USSD application that utilizes web services to provide an interactive disease menu on mobile devices connected to a GSM network, so that they can send notifications to a georeferenced database. A Web GIS module, which obtains spatial data from the database mentioned previously, allows the visualization of potential risk areas, and makes them available in real time to all stakeholders through web-mapping.

The proposed model is provided to governmental organizations, NGOs, and agencies working on health areas, and contains a tool for obtaining georeferenced information on areas or regions affected by some kind of disease while providing decision support for public health professionals. Note that the effectiveness of a GIS system generally depends on the nature of the problem and the selected software, as well as the reliability of the data and the ability of the users to understand their results.

Future work includes an implementation plan that will be drafted for applying the model to a Healthcare institution like the MSPBS in Paraguay. Additionally, a refined cost-comparison model is needed in order to better justify the benefits of the proposed tool.

References

[1] Comisión Económica para América Latina y el Caribe (CEPAL), “TIC y salud: promesas y desafíos para la inclusión social” presented at III Conferencia Ministerial sobre la Sociedad de la Información de América Latina y el Caribe, Lima (Perú), Santiago (Chile), 2010.

[2] Wouters B, Barjis J, Maponya G, Maritz J and Mashiri M. “Supporting home based health care in South African rural communities using USSD technology” presented at Americas Conference on Information Systems (AMCIS), 2009.

[3] Pan American Health Organization (PAHO), World Health Organization (WHO). “New Mobile Phone-based System facilitates Cholera Surveillance and Control in Haiti”. Washington, D.C., 2013. [Online]. Available: http://www.paho.org/ict4health/index.php

[4] Han JH, Sunderland N, Kendall E, Gudes O and Henniker G. “Professional Practice and Innovation: Chronic disease, geographic location and socioeconomic disadvantage as obstacles to equitable access to ehealth”. Health Information Managment *Journal*, vol. 39, no. 2, p.30, 2010.

[5] Lehmann U and Sanders D. “Community health workers: What do we know about them? The state of the evidence on programmes, activities, costs and impact on health outcomes of using community health workers”. School of Public Health, University of the Western Cape. Geneva. 2007.

[6] Direccion General de Estadística, Encuesta y Censo (DGEEC). “Entre los anos 2010 y 2013 se sumaron aproximadamente 780.000 usuarios de la Telefonía Móvil en Paraguay”. 2014. [Online]. Available: http://www.dgeec.gov.py/newspaper/news77.html

[7] Ultima Hora (UH), “El Acceso a Internet de los Hogares”, 2014. [Online]. Available: http://www.ultimahora.com/el-acceso-internet-loshogares-n795345.html

[8] Johnson R. “GIS Technology for Disasters and Emergency Management”. Environmental Systems Research Institute (ESRI). 2000.

[9] Mobicents. “USSD Gateway”. 2008. [Online]. Available: http://www.mobicents.org/

[10] Bertuzzi Leonelli RC and Salazar Fernandez JP. “Diseño e implementación de un servicio de localización y visualización de mapas utilizando J2ME para dispositivos moviles y herramientas de libre distribucion”. Síntesis Tecnológica, 2007.

[11] Siabato Willington,“Métricas aplicadas a los modelos de calidad: caso de uso en los SIG”, Congreso Internacional de Ingeniería Geomática y Topográfica. IX Congreso Nacional Top-Cart, 2008.

Address for correspondence

Silvia Ochoa; nochoa@pol.una.py

Julia Talavera; julia.talavera@pol.una.py

Julio Paciello; julio.paciello@pol.una.py

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-401

Using Publicly Available Data to Characterize Consumers Use of Email to Communicate with Healthcare Providers

Ryan H. Sandefer[a], Saif S. Khairat[b], David S. Pieczkiewicz[a], Stuart M Speedie[a]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[b] *Carolina Health Informatics Program- University of North Carolina, Chapel Hill, NC, USA*

Abstract

The use of patient focused technology has been proclaimed as a means to improve patient satisfaction and improve care outcomes. The Center for Medicaid/Medicare Services, through its EHR Incentive Program, has required eligible hospitals and professionals to send and receive secure messages from patients in order to receive financial incentives and avoid reimbursement penalties. Secure messaging between providers and patients has the potential to improve communication and care outcomes. The purpose of this study was to use National Health Interview Series (NHIS) data to identify the patient characteristics associated with communicating with healthcare providers via email. Individual patient characteristics were analyzed to determine the likelihood of emailing healthcare providers. The use of email for this purpose is associated with educational attainment, having a usual place of receiving healthcare, income, and geography. Publicly available data such as the NHIS may be used to better understand trends in adoption and use of consumer health information technologies.

Keywords:

Health Information Technology; Email; Consumer Involvement.

Introduction

There is increased national attention focused on the use of patient-centered technologies in an effort to engage patients and families in their healthcare, particularly in an effort to address the rise of chronic conditions and managing the complexity of health conditions. The Centers for Medicare and Medicaid Service's (CMS) Electronic Health Record (EHR) Incentive Program and its Meaningful Use objectives have drawn attention to the value of using patient portals and personal health records to better connect patients with their health information. The EHR Incentive Program has four goals: 1) "Improve quality safety, efficiency, and reduce health disparities"; 2) "**Engage patients and families**"; 3) "Improve care coordination, and population and public health"; and 4) Maintain privacy and security of patient health information" [1].

The use of patient portals, personal health records and other patient-focused technologies is intended to improve the level of patient engagement. The Agency for Healthcare Research and Quality (AHRQ) recently completed an Environmental Scan on the topic of patient and family engagement. AHRQ defined patient engagement as "a set of behaviors by patients, family members, and health professionals and a set of organizational policies and procedures that foster both the inclusion of patients family members as active member of the health care team and collaborative partnerships with providers and provider organizations" [2].

A study of US hospitals showed that the adoption of "basic" EHRs has increased by 82% between 2010 and 2011. More importantly, the rates of "comprehensive" EHR use by US hospitals have increased fivefold from 1.6% in 2008 to 8.8% in 2011 [3]. A recent study found similar results regarding patient portals specifically—approximately 50% of US hospitals have patient portals of some kind [4].

There is considerable evidence related to the benefits of using the secure messaging functionality of patient portal systems. In a study involving 35,423 patients with diabetes, hypertension, or both diabetes and hypertension, Zhou, et al. assessed the impact of secure patient-physician email on the quality outcomes as measured by diabetes and hypertension-related quality measures included in the Healthcare Effectiveness Data and Information Set (HEDIS) [5]. The study used administrative data on secure email use between patients and providers, as well as nine HEDIS measures. They reported that "The five leading reasons for patients to email their physicians were to report a change in a condition (16%), discuss lab results (14%), discuss a new condition (12%), discuss changes in prescription dose (11%), and discuss the need for a new prescription (10%)". Physicians receive, on average, between two and twelve messages per day, each requiring an average of 3.5 minutes per response. The findings show that users of secure email had significant improvements in outcomes related to HbA1c screening and control, retinopathy screening, nephropathy screening, and blood pressure control. Interestingly, those patients who had multiple email threads with the provider had an increased likelihood of improved performance on four of the diabetes measures. In other words, the frequency of messaging between a patient and provider is associated with some measures of clinical outcomes improvement.

A similar study focused on provider productivity related to the use of secure messaging tested the impact of the adoption of a patient portal on appointments with physicians, nurse practitioners, and physician assistants in adult primary and urgent care [6]. The study also tested the impact that patient portal adoption had on telephone contact rates with internal medicine and family practice physicians, nurse practitioners, and physician assistants. To evaluate the impact of secure messaging on primary care physicians, the study compared the difference in primary care office visit rates and documented telephone call rates among users and non-users of secure email. In the cohort study, primary care office visits among the sample decreased significantly by 9.7%. Similarly, in a matched control part of the study, the annual office visit rate for those patients using secure messaging decreased by 10.3%, as compared to a 3.7% decrease in the control group. The difference between groups regarding documented telephone

contacts was also significant—controls made 13.7% more telephone contacts than those patients using secure messaging.

Wade-Vuturo, Satterwhite Mayberry [7] conducted a mixed methods study to evaluate the use of a patient portal by patients with type 2 diabetes, secure messaging within the patient portal, and the relationship between secure messaging and glycemic control. The study found that there was a significant association between patients who self-reported frequently using secure messaging within the patient portal—either messaging a provider or requesting an appointment—and lower Hemoglobin A1c values (i.e., glycemic control) [7].

Harris, Haneuse [8] conducted a cross-sectional study to test that secure messaging through a patient portal was associated with higher quality of care and lower care utilization. The patients within the secure messaging group were slightly younger than those in the non-secure messaging group (average age of 58 vs. 63) and were significantly more likely to have a female primary care provider. Regarding the quality measure of HbA1c <7%, patients who had a patient portal account and sent secure messages had higher rates of control than those who had patient portal accounts and did not send messages. In fact, among patients with greater than or equal to 12 threads of secure messages, HbA1c <7% was 36% greater than those in the control group [8].

Despite all of the focus on these patient centered technologies, there has been little growth in the uptake of these technologies among the general population or among those most likely to benefit from its immediate and prolonged use. One major reason for the slow adoption is the slow adoption by healthcare providers and organizations [4]. According to Tenford et al., users of these technologies are most likely to be educated, female and Caucasian [9]. There is not a clear understanding regarding the average age of users, but studies indicate that the majority of users are between 50-70 years of age [10, 11]. Health literacy is cited as a primary barrier to the effective use of personal health records and patient portals for the self-management of chronic disease [12]. According to a study that examined the utilization of patient portals, 88% of Caucasians had a college degree, compared to 49% of African American users [11]. There are also clear ethnic/racial and educational disparities of patient portal use in large integrated delivery systems [11]. While Caucasians are more likely to utilize patient portals, African American patient portal users are significantly more likely to have the presence of a chronic disease [10]. According to one study, "those with limited health literacy were less likely to activate their patient portal account, sign in using their personal login and password, and to use any of the functions" [11].

Many of the major healthcare policy initiatives in the last decade have heavily emphasized patient engagement through the use of technology as a primary objective for improving health outcomes and thereby reducing costs. While the literature is growing on this topic, there is still not adequate knowledge about patients who are using patient portals and how they are using them. Knowledge about use of patient portals by geographic region of the country is minimal. The vast majority of the research on the use of patient-provider email has been within healthcare organizations who were early adopters of health information technology and are overwhelmingly part of integrated delivery networks. There is a need for research on the self-reported use of patient-provider emailing including those with self-reported chronic conditions. The purpose of this study is to use data from the National Health Interview Series, a national, weighted random survey of adults regarding healthcare delivery, to identify the patient characteristics associated with communicating with healthcare providers via email.

Methods

Data Sources and Limitations

This study examined survey data collected as part of the U.S. National Health Interview Series (NHIS) and was recoded, maintained, and distributed by the Integrated Health Interview Series (IHIS) [13]. The NHIS is an in-person, interview data collection program conducted by the Centers for Disease Control and Prevention's National Center for Health Statistics. The purpose of the survey is "to secure accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services rendered for or because of such conditions". The NHIS uses a multi-stage probability sampling to select a random sample of approximately 35,000 households and 87,500 persons. The average response rate for the survey is near 90% of eligible households in the sample. Samples are drawn from each U.S. State and the District of Columbia and it is representative of the U.S. population. The NHIS has been conducted annually since 1957 [14]. Further details about the survey have been published [15, 16]. Data was extracted from the IHIS as a SAS text file, and the SAS text file was read into R and converted into a CSV file. The CSV file was imported into R and converted to a data object for all analysis.

For the purposes of this study, survey data regarding adults ages 18 or greater from 2009, 2011, 2012, and 2013 were combined (Note: the 2010 survey was excluded since it did not address use of email). Between 2009-2013, 492,948 individuals were interviewed across 191,395 households. Analysis of the NHIS data was deemed exempt from review by the University of Minnesota IRB.

The data used in this study included 363,564 individuals 18 years of age or greater. The NHIS includes questions on a variety of health outcomes, demographic variables, and individual characteristics. We used self-reported data on whether the respondent has ever used email to communicate with a healthcare provider. The questionnaire item reads "During the past 12 months, have you ever used computers [to]…communicate with a health care provider by email". Responses were categorized into "Yes", "No", "Refused", "Don't Know", or "Not Ascertained". Only those individuals responding "Yes" or "No" were included in the analysis resulting in a study sample size of 128,100 adults. The following variables were also included in the analyses: sex, gender, race, age, education level, health status, income, geographic region, care access, and diabetes prevalence. Education level was recoded into five categories (less than high school, high school or GED, some college or two year degree, bachelors degree, and graduate degree). Income was recoded into five categories (<35k$, $35-49.9k, $50-74.9k, $75-99.9k, and $100k +). Any response of "Refused", "Don't Know", or "Not Ascertained" were treated as missing values in the analysis.

Analysis

In order to understand the characteristics of adults who email their healthcare providers, proportions, standard errors, and a multiple logistic regression were computed using the R version 3.1.0. Proportions were used to calculate the overall use of email across multiple variables. Logistic regression was used to determine which variables are related to the use of email to communicate with healthcare providers. The following variables were included as independent variables in

the logistic model: sex, gender, race, age, education level, health status, income, geographic region, care access, and diabetes prevalence. The answer to the question about email use was the dependent variable. The results of the logistic regression are presented as odds ratios with the corresponding standard errors 95% confidence intervals, coefficients, standard errors, and Wald Statistic.

Results

The study sample consisted of 128,086 respondents across the four years. The average age of respondents was 37.2 years, and 56% were female. Sixteen percent of respondents had less than a high school diploma, 22% had a high school degree or GED, 30% had some college or a two-year degree, 18% held a bachelor's degree, and 10% held a graduate degree. Table 1 summarizes the survey's respondent characteristics.

Regression results are presented in Table 2. Overall, 5.7% of respondents reported emailing their healthcare provider between 2009 and 2013. Males were significantly less likely to email their provider than females (OR 0.75, CI: 0.71-0.79, $p<0.001$), African American (OR 0.82, CI: 0.76-0.89, $p<0.001$) and Asian (OR 0.85, CI: 0.77-0.93) respondents were significantly less likely to email their healthcare providers than all other races. Age is significantly associated with emailing healthcare providers. Respondents between the ages of 51-60 are significantly less likely to email their providers than those under 30 (OR 0.91, CI: 0.84-1.00, $p<0.05$).

Education level and income were significant predictors of emailing a healthcare provider. The likelihood of emailing a provider is significantly greater for each level of education obtained. Respondents holding a four year college degree significantly more likely to email their provider (OR 11.48, CI: 9.61-13.28, $p<0.001$). Respondents holding a graduate degree were the most likely to email their provider (OR 15.88, CI: 13.25-19.19, $p<0.001$).

Health status is associated with emailing healthcare providers. Adults who report being in excellent health (OR 0.71, CI: 0.59-0.85, $p<0.001$) or good health (OR 0.83, CI: 0.70-1.00, $p<0.01$) are significantly less likely to email their healthcare provider than those who report being in fair or poor health. Patients who report having diabetes are more likely than those without diabetes to email their providers (OR 1.28, CI:1.16-1.41, $p<0.001$).

Geographic region was significantly associated with emailing healthcare providers. Those living in the West were twice as likely to email their provider than those in any other part of the country (OR 2.02, CI:1.88-2.18, $p<0.001$). There was also a significant difference in the likelihood of emailing healthcare providers based upon reporting year. Respondents in 2013 were more likely to email their providers than those responding in 2009 (OR 1.17, CI: 1.10-1.26, $p<0.001$).

Adults who reported having a usual place for obtaining healthcare were also more likely to email their provider than those without a usual place (OR 2.07 CI: 1.88-2.30, $p<0.001$).

Discussion

The study attempted to understand the use of email to communicate with healthcare providers by a nationally representative sample of adults. Because of the nature of the NHIS methodologies, the sample represents the diversity of the United States in terms of age, sex, race, geographic location, and other demographic characteristics. To the best of our knowledge, this is the first study to use a nationally representative sample for assessing email communication between patients and healthcare providers.

Despite increased national attention and increased access to secure email from healthcare organizations, there continues to be a relatively low rate of use of email between adults and their healthcare providers. Overall, 5.6% of adults reported emailing their healthcare provider between 2009 and 2013 (see Figure 1).

Table 1– Email Use with Healthcare Provider by Respondent Characteristics

Characteristics	n = 128,086	Proportion Who Used Email
Male	56,997	4.9%
Female	71,089	6.2%
Diabetes		
No	113,277	5.7%
Yes	12,872	5.1%
Borderline	1,851	5.5%
Age (years)		
18-30	26,505	4.8%
31-40	22,491	7.1%
41-50	21,931	6.6%
51-60	22,104	6.5%
61-70	17,828	6.2%
71+	17,227	2.1%
Education		
Less HS	20,732	0.6%
HighSch	33,271	2.3%
SomeColl	38,819	5.3%
Bachelors	22,312	10.6%
Graduate	12,344	15.9%
Year		
2009	27,437	5.4%
2011	32,695	5.2%
2012	33,887	5.3%
2013	34,067	6.6%
Health Status		
Poor	4,640	3.2%
Fair	14,735	4.0%
Good	75,134	5.7%
Excellent	33,512	6.5%
Income		
<35k$	52,419	2.6%
$35-49.9k	17,416	4.7%
$50-74.9k	19,896	6.7%
$75-99.9k	12,009	8.5%
$100k +	19,094	12.2%
Poverty Level		
In Poverty	21,466	2.0%
Not In Poverty	95,880	6.8%
Geography		
Midwest	27,466	4.6%
Northeast	21,006	5.0%
South	46,714	4.6%
West	32,900	8.5%
Race		
White	96,326	5.9%
Black/AA	19,997	3.7%
AI/AN	1,273	3.1%
Asian	8,004	7.9%
Multi-Race	2,303	6.3%
Usual Place of Care		

Usual Place	54,095	6.2%
No Usual	20,330	2.4%

Table 2– Prevalence of Using Email to Communicate with Healthcare Provider

Object	β	SE β	Wald Stat.	OR (95% CI)
Sex				
Male	-0.29	0.03	-10.86***	0.75 (0.71-0.79)
Diabetes				
Yes	0.25	0.05	5.16***	1.28 (1.16-1.41)
Borderline	0.20	0.11	1.84	1.22 (0.98-1.51)
Age (years)				
31-40	0.04	0.04	0.91	1.04 (0.95-1.13)
41-50	-0.05	0.04	-1.24	0.95 (0.87-1.03)
51-60	-0.09	0.04	-2.05*	0.91 (0.84-1.00)
61-70	-0.13	0.05	-2.67**	0.88 (0.80-0.97)
71+	-0.10	0.07	-14.50***	0.38 (0.33-0.44)
Education				
HighSch	1.10	0.10	11.46***	3.00 (2.50-3.64)
SomeColl	1.81	0.10	19.71***	6.10 (5.11-7.33)
Bachelors	2.44	0.10	26.33***	11.48 (9.61-13.28)
Graduate	2.76	0.10	29.33***	15.88(13.25-19.19)
Year				
2011	-0.10	0.04	-2.50*	0.91 (0.84-0.98)
2012	-0.06	0.04	-1.68	0.94 (0.87-1.01)
2013	0.16	0.04	4.35***	1.17 (1.10-1.26)
Health Status				
Fair	0.02	0.10	0.21	1.02 (0.84-1.25)
Good	-0.18	0.10	-1.97**	0.83 (0.70-1.00)
Excellent	-0.35	0.10	-3.66***	0.71 (0.59-0.85)
Income				
$35-49.9k	0.19	0.05	3.76***	1.21 (1.10-1.34)
$50-74.9k	0.37	0.05	7.94***	1.44 (1.32-1.58)
$75-99.9k	0.48	0.05	9.50***	1.61 (1.46-1.78)
$100k +	0.67	0.05	14.78***	1.95 (1.79-2.14)
Poverty Level				
No Poverty	0.34	0.06	5.54***	1.40 (1.24-1.58)
Geo.				
Northeast	0.00	0.05	0.02	1.00 (0.92-1.09)
South	0.08	0.04	1.10*	1.08 (1.00-1.16)
West	0.70	0.04	18.92***	2.02 (1.88-2.18)
Race				
Black/AA	-0.20	0.04	-4.53***	0.82 (0.76-0.89)
AI/AN	-0.46	0.17	-2.76**	0.63 (0.45-0.86)
Asian	-0.17	0.05	-3.50***	0.85 (0.77-0.93)
Multi-Race	0.02	0.10	0.21	1.02 (0.85-1.22)
Usual Place of Care				
Usual Place	0.73	0.05	13.98***	2.07 (1.88-2.30)

CI – confidence interval; OR =odds ratio; SE β = standard error of β coefficient; Goodness of Fit = X^2 = 114,039 p=1.0
* Significance of Wald Statistic at 0.05
** Significance of Wald Statistic at 0.01
*** Significance of Wald Statistic at 0.001

Many of the findings from the current research are consistent with previous research—age, education level, income, and race are associated with emailing healthcare providers[10, 11, 17]. Previous research has indicated that the most significant users of secure email are between 50-70 years of age, but our research indicates that younger individuals under the age of forty are somewhat more likely to email their healthcare provider.

Of particular interest for this research and potentially related to the impact of the CMS EHR Inventive Program is that respondents in 2013 were significantly more likely to email their healthcare providers than those in 2009 (5.4% in 2009 and 6.6% in 2013). Females increased their use even more significantly over the four years (5.6% in 2009 to 7.4% in 2013). The level of EHR adoption has increased dramatically in both outpatient and inpatient areas since 2008 likely due to the federal EHR Incentive Program, particularly the program's requirement for healthcare organizations to provide patients the means to send and receive secure email messages from their healthcare providers.

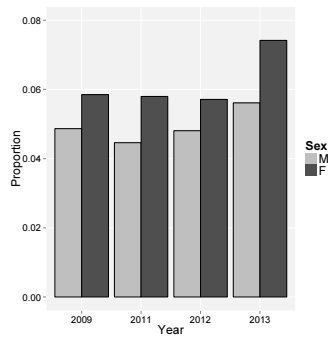


Figure 1– Proportion of Respondents Emailing Healthcare Providers By Year and Sex

Education level is a major factor related to emailing healthcare providers. Educational attainment has been shown to be positively related to overall "health literacy" [18], and adults with lower levels of health literacy are less likely to use internet based patient portals to communicate with healthcare providers [11]. Our findings indicate that those with lower education levels risk falling behind as healthcare organizations increasingly turn to health information technology for coordinating healthcare services, providing health information to patients, and communicating with patient about their care.

Moreover, research has shown that individuals with lower health literacy are less likely to use health information technology [7, 11] and individuals who use health information technology have improved clinical outcomes [8]. Our research indicates that health status is significantly associated with using email to communicate with healthcare providers. Those individuals who report being in fair or poor health are significantly more likely to email their healthcare provider than those who report good or excellent health. Future research should investigate the relationship between health status and education level on reported email use with healthcare providers.

Another interesting finding relates to healthcare access. Those adults who reported having a regular place of care were significantly more likely to email their healthcare providers than those who do not. As the Affordable Care Act increases access to the healthcare delivery system for millions of Americans, these individuals with increased access to

healthcare services could potentially have an impact on the use of email communication with healthcare providers.

This study has several limitations that need to be taken into consideration when reviewing these results. It is not clear from the data what the reasons are for emailing healthcare providers or how the email communication influenced the healthcare encounter. The data are also limited in that email is a self reported measure and does not provide the frequency of emailing providers. Future research should also consider investigating the multiple relationships among socio-demographic characteristics, patient-provider email, and clinical outcomes.

Conclusion

Healthcare stakeholders are increasingly aware that technology is a critical factor for monitoring chronic conditions, improving decision-making, and creating a healthcare system that is more proactive and less reactive. As the United States healthcare system continues to move additional services online and expect patients to interact with their care providers via the Internet, those individuals already suffering from health disparities risk falling further behind.

The use of patient-centered technologies to improve chronic disease management is increasingly important [19, 20]. Generally, these technologies have had limited functionality, with much of the capabilities revolving around entering or reviewing health data. The technologies, however, are increasingly intelligent and interactive, thereby providing both patients and clinicians with triggers regarding for example interactions of drugs or allergies that could lead to an abnormal medical event [21]. As the HIT industry continues to advance these patient-centered tools, this study's findings regarding the frequency of such email use and the predictors of email use between U.S. healthcare providers and adults should be considered when developing tools, developing educational materials, and promoting the use of these new tools to patients.

References

[1] CMS. Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule. Federal Register 2012: 77(171): 53967-4162.

[2] AHRQ. Patient and Family Engagement: Environmental Scan Report Rockville, MD: Agency for Healthcare Research and Quality; 2012. Available from: http://www.ahrq.gov/research/findings/final-reports/ptfamilyscan/ptfamily1.html.

[3] Charles D, Furukawa M, Hufstader M. Electronic Health Record Systems and Intent to Attest to Meaningful Use among Non-federal Acute Care Hospitals in the United States: 2008-2011 Washington, D.C.: Office of the National Coordinator for Health Information Technology, 2012, Contract No.: 1.

[4] Terry K. Patient Portal Explosion Has Major Health Care Implications: iHealthBeat; 2013 [cited 2014 February 1]. Available from: http://www.ihealthbeat.org/insight/2013/patient-portal-explosion-has-major-health-care-implications.

[5] Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Affairs 2010: 29(7): 1370-5.

[6] Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. American Journal of Managed Care 2007: 13(7): 418-24.

[7] Wade-Vuturo AE, Satterwhite Mayberry L, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. Journal of the American Medical Informatics Association 2013: 20(3): 519.

[8] Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009: 32(7): 1182-7.

[9] Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. Journal Of General Internal Medicine 2012: 27(4): 420-4.

[10] Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. Journal of the American Medical Informatics Association: JAMIA 2011: 18 Suppl 1:i8-12.

[11] Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The Literacy Divide: Health Literacy and the Use of an Internet-Based Patient Portal in an Integrated Health System—Results from the Diabetes Study of Northern California (DISTANCE). Journal of Health Communication 2010: 15: 183-96.

[12] Mitchell B, Begoray DL. Electronic personal health records that promote self-management in chronic illness. Online Journal of Issues in Nursing 2010: 15(3).

[13] Minnesota Population Center and State Health Access Data Assistance Center Minneapolis: University of Minnesota; 2012. Version 5.0:[Integrated Health Interview Series].

[14] NHIS. National Health Interview Survey: CDC/National Center for Health Statistics; 2012 [cited 2014 February 2]. Available from: http://www.cdc.gov/nchs/nhis/about_nhis.htm.

[15] Johnson PJ, Blewett LA, Ruggles S, Davern ME, King ML. Four decades of population health data: the integrated health interview series as an epidemiologic resource. Epidemiology 2008: 19(6): 872-5.

[16] Iezzoni LI, Yu J, Wint AJ, Smeltzer SC, Ecker JL. Conditions causing disability and current pregnancy among US women with chronic physical disabilities. Medical care 2014: 52(1): 20-5.

[17] Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. Journal Of General Internal Medicine 2011: 26(10): 1117-23.

[18] Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Archives of internal medicine 1998: 158(2): 166-72.

[19] Bates DW, Bitton A. The future of health information technology in the patient-centered medical home. Health Affairs 2010: 29(4): 614-21.

[20] Hannan A. Providing patients online access to their primary care computerised medical records: a case study of sharing and caring. Informatics in Primary Care 2010: 18(1): 41-9.

[21] Luo G. Triggers and monitoring in intelligent personal health record. Journal Of Medical Systems 2012: 36(5): 2993-3009.

Address for correspondence

Ryan Sandefer

Sande719@umn.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-406

Health Information Technology Evaluation Framework (HITREF) Comprehensiveness as Assessed in Electronic Point-of-Care Documentation Systems Evaluations

Paulina S. Sockolow[a], Kathryn H. Bowles[b], Michelle Rogers[c]

[a] *Drexel University College of Nursing and Health Professions*
[b] *University of Pennsylvania School of Nursing*
[c] *Drexel University College of Computing and Informatics*

Abstract

We assessed the Health Information Technology (HIT) Reference-based Evaluation Framework (HITREF) comprehensiveness in two HIT evaluations in settings different from that in which the HITREF was developed. Clinician satisfaction themes that emerged from clinician interviews in the home care and the hospital studies were compared to the framework components. Across both studies, respondents commented on 12 of the 20 HITREF components within 5 of the 6 HITREF concepts. No new components emerged that were missing from the HITREF providing evidence that the HITREF is a comprehensive framework. HITREF use in a range of HIT evaluations by researchers new to the HITREF demonstrates that it can be used as intended. Therefore, we continue to recommend the HITREF as a comprehensive, research-based HIT evaluation framework to increase the capacity of informatics evaluators' use of best practice and evidence-based practice to support the credibility of their findings for fulfilling the purpose of program evaluation.

Keywords:
Evaluation studies; Technology evaluation; Health information systems; Medical informatics.

Introduction

The increase in health information technology (HIT) implementation worldwide has created a need to better ensure the realization of the system's intended benefits using rigorous evaluation methods. Informatics evaluators tend to use frameworks, that is, models that describe the interrelationships among variables [1], describe the relationship between a framework dimension and a result [2, 3], or describe an implementation model [4]. However, most frameworks do not include the three contextual aspects – organizational, systemic and environmental (political), and professional – that influence whether or how the system will be used. Furthermore, frameworks with social and organizational evaluation dimensions lack clarity or specificity of their evaluation criteria, or have inadequate focus to assess both individual and organizational aspects of evaluation. For example the HOT-fit framework, which is theory-driven, does not include the professional contextual aspect [5].

To address the three-contextual-aspects deficit, the author (PS) developed and assessed the HIT Reference-based Evaluation Framework (HITREF) in the evaluation of an EHR implemented in a geriatric day care center[5, 6]. The framework included health services research evaluation methodologies to extend the informatics evaluators' focus beyond user, technical, and organization interactions to include organizational contextual considerations and other stakeholders' perspectives. The HITREF is a comprehensive HIT evaluation framework firmly grounded in research evidence that identifies a range of clinician satisfaction characteristics and dimensions to be measured. This framework was the result of a comprehensive literature review of over 2000 HIT evaluation studies [5] that expanded on the work of Ammenwerth and de Keizer who reviewed 15,000 articles [7]. The HITREF provides a comprehensive list of 22 criteria related to clinician satisfaction with HIT as themes for the study analyses, as shown in Figure 1. The criteria are organized along the following axes: Structure/Logistics/Process/Outcome with two additional concepts: *Barriers/Facilitators to Adoption* and *Unintended Consequences/Benefits*. Three components added the organizational context: Diffusion; HIT Selection/ Development/Implementation/Training; and Unintended Consequences/Benefits. Also Barriers/Facilitators to Adoption added systemic and environmental (political) contexts. Lastly, three components added the professional context (e.g., patient perspective): Patient Privacy; Patient Satisfaction with EHR; and Patient Satisfaction with Care [5].

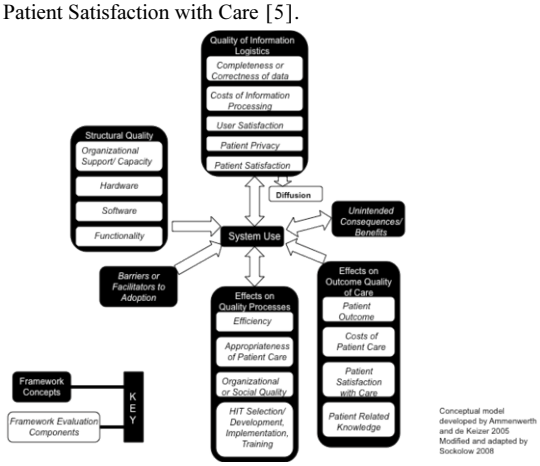


Figure 1 – The Health Information Technology Reference-based Evaluation Framework

This manuscript describes further assessment of HITREF comprehensiveness when used in evaluations of HITs in settings different from that in which the HITREF was originally developed. The new systems evaluated were point-of-care documentation systems used by health care professionals on multi-disciplinary teams who delivered direct patient care: a home care agency electronic health record (EHR) and a

hospital nursing information system (NIS).

Methods

To assess HITREF comprehensiveness, the clinician satisfaction themes that emerged from the home care and the hospital studies were compared to the framework. The home care study used an embedded mixed methods design to collect and analyze actual system usage, and post intervention clinician interviews and survey responses to assess satisfaction with the system [8]. The hospital study, carried out in two hospitals, employed scenario testing with qualitative analysis to assess clinician usage and satisfaction. The scenario study design was presented as a modified think-aloud protocol [9]. This protocol is a standard methodology used to elicit data about cognitive reasoning that occurs during a problem solving task. Twelve users were presented with typical usage scenarios and allowed to walk through how they would complete the action requested using the NIS [10]. Institutional Review Boards approved the studies.

Both the hospital and home care sites had implemented commercially available point-of-care documentation systems in which clinicians recorded care plans, interventions, and outcomes. The EHR was implemented in the home care site in 2009. It supported documentation management and had limited interoperability with health system hospitals. The study was conducted from 2008 to 2011 [8]. The hospitals implemented an NIS in 2011 that functioned with the Computerized Provider Order Entry system within the EHR. Nurses selected from the NIS's approximately 200 interdisciplinary evidence-based clinical practice guidelines to guide and document patient care. The NIS was evaluated in 2012 [10].

In the home care study (PS), semi-structured interviews were conducted until saturation eliciting information about clinicians' areas of concern or satisfaction with the EHR. Content analysis of interview responses started with the HITREF and was followed by mapping the coded themes to the framework, creating a conceptualization that encompassed all participants' experiences [8]. In the hospital study (KB,PS,MR), a convenience sample of consented nurses from selected floors participated in the scenario testing until saturation was reached. Satisfaction with NIS was assessed using scenario testing, which entailed the researchers' presenting to participants previously prepared scenarios and follow-up interview questions while observing their NIS use in a conference room on the unit. For each audio-recorded transcript, researchers coded themes independently and then coded together in relation to the HITREF. A 10% double reliability check was performed. Researchers retained any theme that did not match HITREF criteria (i.e., component within a larger HITREF concept) to add it to the framework as a new HITREF component. An unmatched theme was of special interest as it indicated an area where the framework was not comprehensive [10].

Results

Reported here are the results from both the home health agency and the hospitals studies related to HITREF comprehensiveness. The home health study included analysis of survey responses from 71 clinicians (52% of eligible participants) as well as observations and interviews with 6 (4%) clinicians [8]. The hospital NIS study involved 12 nurses on 2 units who participated in the scenario testing [10].

Home health clinicians were dissatisfied with initial and ongoing training and field support. They were satisfied with the hardware availability and dissatisfied with frequent system problems, as well as with EHR software related to usability and functionality. Overall clinicians were satisfied with the EHR data completeness and timeliness, which improved data availability and supported providing care and team communication. Some clinicians were dissatisfied with the unintended consequence of problems using the EHR, which negatively impacted patient care. Clinicians were also dissatisfied with EHR impact on efficiency resulting from a mismatch between the task requirements and the software functionality. While the EHR was observed to improve accessibility to clinical information, clinicians were dissatisfied with EHR impact on appropriateness of care. Clinicians perceived that the computer disrupted their establishment of patient rapport thereby impacting patient adherence. Clinicians were mostly satisfied with EHR impact on organizational/ social quality related to team communication. On the other hand, clinicians were particularly dissatisfied with their perceived lack of involvement in system selection/ development/ implementation/ training. Lastly, clinicians identified a barrier to adoption: inadequate interoperability as exemplified by their inability to access laboratory results from sources external to the health system [8].

In summary, home health clinicians commented on the following 12 HITREF components, organized within 5 *HITREF concepts* (examples are provided in Table 1):

- *Structural Quality*: Organizational support/capacity, Hardware (e.g., system availability [5]), Software (e.g., usability [5]), Functionality (e.g., tools and resources [5]);
- *Quality of Information Logistics*: Completeness/correctness of data (e.g., data quality [5]), User satisfaction;
- *Unintended Consequences* (e.g., benefits or adverse results [5]); and
- *Effects on Quality Processes*: Efficiency (e.g., time required for tasks [5]), Appropriateness of Patient Care (e.g., "medical efficiency" such as adherence to protocols [5]), Organizational or social quality (e.g., cooperation or communication [5]), HIT Selection/ Development/Implementation/Training; and
- *Barriers or Facilitators to Adoption* (e.g., perceptions related to implementations [5]).

The hospital NIS study findings indicated that the participating nurses universally preferred documenting in the NIS rather than return to paper records. As seen in the home health study, hospital nurses were dissatisfied with the ongoing HIT training as well as software usability and functionality. They were dissatisfied with computer placement, which required standing while using the computer. Also as seen in the home health study, nurses were satisfied with the completeness and timeliness of the documentation. Nurses were dissatisfied with the changes introduced by the NIS, which included redundant documentation and bottlenecks, thereby reducing efficiency. They were also dissatisfied with NIS impact on appropriateness of care related to increased time at the bedside documenting and decreased time spent providing direct patient care. Related to Organizational/social quality, the potential to realize team communication among the clinical roles and disciplines was not always realized as evidenced by duplicate documentation and redundant questions posed to the patient [10].

The hospital nurses commented on the following 8 HITREF components within 3 *HITREF concepts*:

- *Structural Quality*: Organizational support/capacity, Hardware, Software;
- *Quality of Information Logistics*: Completeness/correctness of data, User satisfaction; and

- *Effects on Quality Processes*: Efficiency, Appropriateness of Patient Care, Organizational/social quality (team communication).

Across both studies, respondents commented on 12 HITREF components across 5 HITREF concepts. All themes that emerged from either study matched a HITREF component.

Table 1 - Examples of HITREF Concepts/Components Occurring in Each Study

HITREF Concept	**HITREF Component**	**Home Care**	**Hospital**
Structural Quality	Organizational Support/capacity	-Field Support, Training	- Training; Communication
	Hardware	- Reliability	- Computer placement
	Software	- Usability: Navigation	- Usability: Navigation; Mismatch screenflow/ workflow; Documentation fatigue
	Functionality	+Memory prompts	-Absence of needed functionality
Quality of Information Logistics	Completeness/ correctness of data	+ Timely, complete	+ Accessible, Complete
	Costs of information processing	NA	NA
	User satisfaction	+Overall satisfaction	+Overall satisfaction
	Patient privacy	NA	NA
	Patient satisfaction with EHR	NA	NA
	Diffusion	NA	NA
Unintended Consequences/Benefits	Unintended Consequences/Benefits	- EHR problems interfere with patient care	NA
Effects on Outcome Quality of Care	Patient outcome	NA	NA
	Costs of patient care	NA	NA
	Patient satisfaction with care	NA	NA
	Patient related knowledge	NA	NA
Effects on Quality Processes	Efficiency	- Increased documentation	- Bottlenecks; Redundant documentation; Many checkboxes
	Appropriateness of patient care	-Patient rapport	+ Time at bedside - Patient time, more NIS time
	Organizational or social quality	+Team communication	-Potential for increased multidisciplinary team communication not realized
	HIT Selection/ Development/ Implementation/ Training	- Clinician involvement	NA
Barriers or Facilitators to Adoption	Barriers or Facilitators to Adoption	- Interoperability	NA

Note: + indicates satisfactory attributes; - indicates dissatisfactory attributes; NA indicates not applicable

Discussion

The HITREF was designed to be used to evaluate a broad range of HIT. To date, the framework has been used to assess two types of HIT in three different settings, [6, 8, 10]. In addition to being evidence-based, a HITREF advantage is the inclusion of organizational, systematic and environmental, and professional criteria to evaluate whether the HIT was used as intended [5]. Across the two evaluations described in this manuscript, [8, 10] these three levels of criteria had one or more components with matched themes (e.g., Organizational support/capacity, Barriers/facilitators to adoption, User satisfaction, respectively). The emergence of these themes underscores the HITREF's value as an evaluation tool that includes contextual concerns in addition to technological capabilities. The initial study in which the HITREF was assessed [5] and the two studies described in this manuscript [8, 10] were similar in the inclusion and exclusion of HITREF components in the evaluations, despite differences in settings and HIT evaluated. The absence of new components missing from the HITREF indicates that the HITREF is a comprehensive evaluation framework.

Overall, 60% of the HITREF components emerged in the two recent studies across all but one of the HITREF concepts, as shown in Table 1. The excluded concept, *Effects on Outcome Quality of Care*, contains four patient-focused components without related themes that emerged in the evaluations. The constituent component, Costs of patient care, is an issue not usually encountered by front line clinicians. Two components that address the patient perspective (i.e., Patient satisfaction with care; Patient related knowledge) may be more likely to emerge in evaluations of patient-facing systems such as personal health records, EHR patient portals, or mobile health applications. Notable is the presence of patient outcomes in this group of unmatched components. Possibly clinicians did not see a relationship between their use of the point-of-care documentation system and the organization's goal of system implementation to improve patient outcomes. Similarly, the component Diffusion in the *Quality of Information Logistics* concept has a focus on an issue not usually encountered by front line clinicians: whether the system is universally used. Two additional components in the *Effects on Outcome Quality of Care* concept that addressed the patient perspective (i.e., Patient privacy concerns; Patient satisfaction with EHR) did not have themes emerge, probably because the studies did not assess patient-facing systems.

The 60% component match and lack of new components support the applicability of HITREF in evaluations with a range of HIT, settings, and users. Although the framework was

drawn from literature mostly focused on physician HIT use in hospitals [5, 7], the studies involved nurses in hospitals and multi-disciplinary teams in community settings, and focused on both EHRs and NIS. However, it is possible that HITREF use in evaluations of other HIT may identify new components to be added to the HITREF.

The two evaluations also provided support for another aspect of HITREF applicability: its use by informatics evaluators. While the first author (PS) developed the HITREF and conducted the studies, other members of the research team (KB, MR) successfully used the HITREF in their analyses (8, 10]. Use by informatics researchers other than the HITREF developer is evidence that the HITREF belongs in informatics evaluators' toolboxes.

As suggested in our initial study, because the HITREF is relatively large, the evaluator may choose to select HITREF components related to the questions being asked. For example, an evaluator could include questions about patient care and efficiency in evaluations of HIT intended to improve the clinical process or exclude questions about patient perspective from a point-of-care documentation system study.

A research opportunity previously identified that is yet to be addressed is increased knowledge about the relationships among the HITREF evaluation components. This exploration will be possible by using the HITREF in evaluation studies of larger sample sizes.

Conclusion

The HITREF fulfills the purpose of a framework: to ensure a study is comprehensive in terms of inclusion of stakeholder concerns to promote engagement and obtaining full information, and improve decision-making. HITREF use in evaluations with a range of HIT, settings, and users, by researchers new to the HITREF, demonstrates that it can be used as intended. Therefore, we continue to recommend the HITREF as a comprehensive, research-based HIT evaluation framework to increase the capacity of informatics evaluators' use of best practice and evidence-based practice to support the credibility of their findings for fulfilling the purpose of program evaluation.

Acknowledgements

Agency for Healthcare Research and Quality Exploratory Grant 1R21HS021008-01; Drexel University Career Development Grant; National Library of Medicine/ Robert Wood Johnson Foundation Public Health Informatics Training Grant T15LM07452; The Johns Hopkins University School of Medicine. Barbara Granger for editorial expertise.

References

[1] Davis FD, Bagozzi RP, and Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989: 35(8): 982.

[2] Labkoff SE, and Yasnoff WA. A framework for systematic evaluation of health information infrastructure progress in communities. J Biomed Inform 2007 4: 40(2): 100-5.

[3] Sittig DF, Shiffman RN, Leonard K, Friedman C, Rudolph B, Hripcsak G, et al. A draft framework for measuring progress towards the development of a National Health Information Infrastructure. BMC Med Inform Decis Mak 2005: 5(1): 14.

[4] Stead WW, Haynes RB, Fuller S, Friedman CP, Travis LE, Beck JR, et al. Designing medical informatics research and library--resource projects to increase what is learned. J Am Med Inform Assoc 1994: 1(1): 28-33.

[5] Sockolow PS, Crawford PR, and Lehmann HP. Broadening a general framework for evaluating health information technology through health services research evaluation principles. Methods of Medical Information 2012: 51(2): 122-130.

[6] Sockolow PS, Weiner JP, Bowles KH, Abbott P, and Lehmann HP. Advice for decision makers based on an electronic health record evaluation at a Program for All-inclusive Care for Elders site. Appl Clin Inf 2011: 2(1): 18-38.

[7] Ammenwerth E, and de Keizer N. An inventory of evaluation studies of information technology in health care - Trends in evaluation research 1982-2002. Methods Inf Med 2005: 44(1): 44-56.

[8] Sockolow PS, Bowles KH, Adelsberger MC, Chittams JL, and Liao C. Challenges and facilitators to adoption of a point-of-care electronic health record in home care. Home Health Care Services Quarterly 2013: 192: 939.

[9] Nielsen J. Usability Engineering. San Diego, CA: Academic Press, 1993.

[10] Sockolow PS, Rogers M, Bowles KH, Hand KE, and George J. Challenges and facilitators to nurse use of a guideline-based nursing information system: Recommendations for nurse executives. Appl Nursing Res 2013: 27(1): 25-32.

Address for correspondence
Paulina Sockolow DrPH MS MBA
College of Nursing and Health Professions
Drexel University
Mail Stop 71044
7th Floor, Room 706
1601 Cherry Street, MS 71044
Philadelphia, PA 19102
pss44@drexel.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-410

Patient Outcomes as Transformative Mechanisms to Bring Health Information Technology Industry and Research Informatics Closer Together

Jacob Krive[a,b,c]

[a] *Department of Information Systems, Advocate Health Care, Oak Brook, IL, USA*
[b] *Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, USA*
[c] *Department of Biomedical Informatics, Nova Southeastern University, Fort Lauderdale, FL, USA*

Abstract

Despite the fast pace of recent innovation within the health information technology and research informatics domains, there remains a large gap between research and academia, while interest in translating research innovations into implementations in the patient care settings is lacking. This is due to absence of common outcomes and performance measurement targets, with health information technology industry employing financial and operational measures and academia focusing on patient outcome concerns. The paper introduces methodology for and roadmap to introduction of common objectives as a way to encourage better collaboration between industry and academia using patient outcomes as a composite measure of demonstrated success from health information systems investments. Along the way, the concept of economics of health informatics, or "infonomics," is introduced to define a new way of mapping future technology investments in accordance with projected clinical impact.

Keywords:

Economics of health information systems; health information technology; research informatics; health outcomes research; integration of academia and industry.

Introduction

As relatively young niches of information technology (IT) in the process of maturation, health information management (HIM) and health information technology (HIT) have gone through several stages of (first) chaotic and (later) systematic development. In some countries such as the United States, much of the growth is encouraged by government subsidies and legislative changes penalizing healthcare organizations for failing to invest in HIT. Early skepticism regarding benefits of digitization of medical records and computer-assisted clinical workflow turned into excitement regarding HIT's ability to drive quality improvement efforts resulting from electronic medical records (EMR) installations and streamlining ordering process via computerized physician order entry (CPOE) applications. In response to subsidies, large metropolitan healthcare systems in the United States championed HIT investment. The rate of adoption was closely associated with large size, urban location, and health maintenance organizations (HMO) penetration [1]. HMOs rely, in part, on health outcomes of the populations they serve in order to remain financially solvent. Therefore, association of HIT with quality of patient care puts technology investment on the central stage for organizations interested in generating income through advances in health outcomes.

Early documented and perceived gains from EMR and CPOE systems in the processes of clinical workflow enhancement, quality improvement, and reduction of medical errors, led to further development of disciplines that relied on the marriage of information and medical sciences. Examples of such disciplines are "big data" analytics, population health, predictive analytics, clinical decision support systems (CDSS), natural language processing, mobile health, health information exchanges, and even social engineering. In the United States, EMR adoption was systematized and documented in several widely referenced models, such as HIMSS EMR Adoption Model [2] and the Meaningful Use provision of the Accountable Care Act. The latter model is a government legislation that mandates interoperability of disparate HIT applications.

Despite massive investments into HIT by healthcare organizations and software vendors worldwide, as well as significant growth in the biomedical informatics research and education in academia, there remains a large gap between industry and academia. Moreover, some of the massive commercial investments into HIT, fueled by industry standards modeling and government subsidies, have yet to provide direct link between technology growth and actual patient outcomes. Seemingly, healthcare organizations, software companies, and universities are interested in similar goals – error prevention, quality improvement, enhanced clinical workflows, predictive analytics, and breakthroughs in knowledge processing and management. In reality, industry focused on digitizing medical records and streamlining orders (EMR/CPOE), while academia emphasized new methods of disease management and surveillance, as well as specific applications of medical knowledge via informatics for targeted disease treatment. This paper is aimed at exploring shared goals potentially leading to transformation of HIT and closer collaboration between industry and academia to achieve common objectives.

Patient Outcomes Focus as a Common Goal for Industry and Academia

On the industry side, transition from paper records to EMR and subsequent automation of clinical workflows and communication processes via CPOE and unified communications achieved a plateau in the past few years. At this level, a majority of the most successful organizations have already implemented these technologies and look to capitalize upon them in order to achieve more advanced clinical objectives. These massive investments travel through the process that Gartner calls the Hype Cycle – transitioning

through trigger, peak, disillusionment, enlightenment, and plateau of productivity [3]. As HIT travels through the hype cycle, and billions of dollars are invested into it, it becomes apparent that the HIT outcomes measurement process needs to be centered around clinical outcomes, especially in the absence of real productivity gains [4] expected of initial technology investments. While technologies such as CDSS have demonstrated success in driving effectiveness of specific process measures [5], HIT research remains in the early stage of demonstrating measurable clinical impact [6].

It is apparent that technology by itself will not positively impact safety. Instead, technology supported clinical workflow changes could hold the biggest promise of making measurable impact on quality and safety of patient care [7]. Massive and frequently chaotic investment in HIT needs to be replaced by targeted clinical problem definition that initiates the process of technology investment [8], followed by routinization – the process by which using innovation becomes regular organizational practice [9]. But why does the process of routinization remain so painfully slow in transition from biomedical informatics to HIT? It is due to mismatched effectiveness evaluation measures: while HIT focuses on operational measures and return on investment (ROI) from infrastructure and applications, the research side emphasizes patient outcomes over cost and operations.

The Evolving Role of HIT and Its Effectiveness Evaluation Criteria

Biomedical informatics research of health outcomes resulting from information technology development has remained separate from "field" investments at hospitals and medical practices that were largely driven by either government spending and/or perceived expectations of quality and productivity gains. As healthcare cost pressures rise and the healthcare delivery model shifts from volume to value based reimbursement [10], physician to patient centered care (Figure 1) and physician oriented to consumer oriented communication model (Figure 2), simply running an efficient information technology organization is not satisfactory. Providing customers with appropriate technology services, fitting within budget, maintaining up-to-date infrastructure, and timely completion of projects become secondary measures of HIT performance.

Figure 1 – Patient centered care model

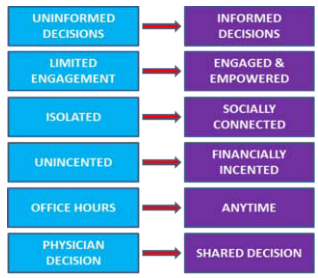


Figure 2 – Consumer oriented care model

A common denominator rises between HIT and research informatics, and that is health outcomes orientation. The primary goal of every healthcare organization is patient services, and so every process needs to be streamlined and oriented at achieving patient satisfaction. In turn, healthcare organizations' ability to deliver the most effective care at the lowest cost with the best clinical outcomes is a must in the new business of medicine environment where health outcomes and reimbursement are tied together. This applies to healthcare models in both private and public sectors, as ultimately no matter who controls reimbursement, the goals of achieving effective care at the lowest cost are the same. Patient outcomes orientation offers industry and academia a unique common perspective that could spawn new collaboration projects aimed at increasing the pace of translating discovery into industry implementation.

Health outcomes focus leads to emergence of a new discipline – let's call it economics of health informatics, or simply "health infonomics." The health infonomics science would merge biomedical informatics research with health economics and HIT investment mechanisms to generate new models of targeted selection and measurement of information technology aimed at reaching specific clinical and productivity goals. Health infonomics would also help construct instruments of innovation and operational effectiveness, in terms of impact on clinical outcomes.

The Evolving Model of HIT Effectiveness and Performance Evaluation

Changes in perception, evaluation, and measurement of the effectiveness of healthcare information systems are likely to lead a change in expectations and assessment of IT operations in healthcare organizations. Three traditional measures of IT effectiveness – service, quality, and cost – will be tied to patient-centric outcomes (Figure 3). This structure mimics industry service model change from physician-centric to patient-centric processes aimed at delivering accessible high-quality care.

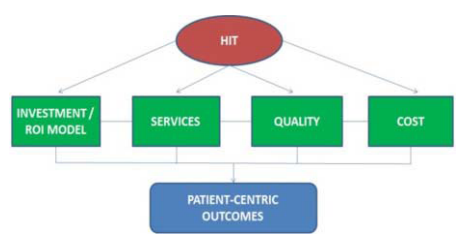


Figure 3 – Patient centric orientation of HIT

In accordance with this patient-centric care impact measure, future HIT investment, as well as operational effectiveness and customer satisfaction, will require a new inclusive measure of success represented by combination of clinical, technology, and cost factors. Patient Care Index (Figure 4), aimed at measuring the overall effectiveness of HIT investment and operations, would link technology to patient outcomes, removing separation between technology acquisition and enterprise clinical goals of the healthcare organizations. Such fundamental shift to tying success of information technology to clinical outcomes would drive re-evaluation of existing ROI models to tie them to core clinical and cost objectives of healthcare organizations.

Figure 4 – Patient care index as a central measure of HIT success

The patient care index (Tables 1 and 2) is a composite statistical score that includes a multitude of factors ranging from clinical outcomes, directly linked to specific technologies, to operational effectiveness in terms of systems stability and availability to customer satisfaction scores. Factors would be ranked and assigned statistical weights in accordance with clinical impact and the potential of HIT to impact patient and/or hospital operations outcomes. Factors listed in the tables below are organized by clinical and non-clinical items.

Table 1 – Example of the patient care index factors: Clinical

Health Outcomes	Quality	Safety
Mortality	Core measures compliance	Unassisted falls percentile
Preventable readmissions	Corrected reports	Orders placed via CPOE sets
Length of stay	Ambulatory patient lab redraws	Transfusion safety
ICU ventilator days	Acute care hospitalization rate	Infection control chart
NICU mortality index	Management of oral medications rate	Adverse drug effects (ADE) rate
Pediatric surgery mortality index	Improvement in dyspnea rate	Lab identification errors
	Improvement in ambulation rate	Mislabeled pathology slides
	CPOE utilization rate	Serious safety event rate

Table 2 – Example of the patient care index factors: Financial, operational, and technology

Cost	Patient Engagement	Technology
Cost per inpatient discharge	Patient satisfaction survey rate	Total critical care systems uptime
Cost per outpatient visit	Percent of (online) socially engaged patients	Total financial systems uptime
Total ED visits	Number of patient portal users	Number of downtime events
Total outpatient visits	Number of unique patient portal logins	Associate and provider IT services satisfaction score
Cost per ED visit	Health and wellness index	IT operating cost
Inpatient admissions	Public Wi-Fi uptime score	CPOE adoption rate

Project management and selection of computer applications to meet the needs of a clinical enterprise would undergo changes in accordance with health outcomes orientation of the technology selection process. Instead of identifying technological needs and subsequently evaluating technology's potential to bring desired benefits to the organization, the process would reverse to identification of a specific clinical problem – leading to a technology solution (Figure 5). Should there be no existing solution to a clinical problem, it is potential for collaboration with academia, leading to not only breakthrough innovation but also quick translation to industry oriented solution and implementation in patient care settings, using common goals between HIT and academia models.

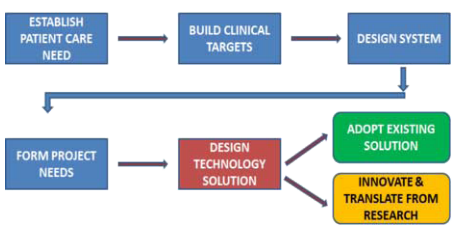


Figure 5 – Health outcomes oriented project selection and management

Infonomics, Health Outcomes, and Research/HIT Collaboration

The first two steps in the health outcomes oriented project management model described in Figure 5 - establishing patient care need and building clinical targets – are rooted in evidence based clinical practices. The figure can be adjusted based on clinical specialty and target patient population. For example, geriatric medicine would pose unique challenges and needs for patients who require physical and cognitive interventions aimed at improving and maintaining their health [11]. Establishing patient care need would represent a

complex process of determining the most critical factors affecting geriatric patients and subsequent opportunities for computer-assisted interventions. The outcomes based project management process would require detailed investigation of clinical needs, followed by setting quantitative targets for potential HIT interventions, and analysis of the current and potential options for implementation. This model brings HIT practitioners and researchers together to solve clinical problems. In another example, a post facto study investigated effectiveness of pneumonia CPOE order sets measured by such health outcomes as mortality, readmission, comorbidity, and length of stay [12]. The patient outcomes tied technology selection and management model would require preliminary investigation of expected health outcomes prior to system implementation and performing post-implementation analysis of the outcomes to determine clinical effectiveness of the HIT initiative.

Conclusion

Reevaluation of HIT goals, expectations, and measurement processes would initiate changes in the healthcare technology market, as well as investment selection and allocation processes at the healthcare provider organizations. Ultimately, providers, patients, and suppliers are expected to benefit due to renewed focus on patient care quality and targeted healthcare technology spending. Such collaboration could also have a profound positive impact on cost, as a secondary measure of industry/academia partnership success. But ultimately patients and providers would benefit most from integration of goals that could help bring industry and academia closer together in search of common scientific outcomes in line with business objectives. Academia is already focused on patient outcomes and has been for a long time. Industry's profit and operations models included a large variety of performance measures, not all leading to improved patient outcomes. Realignment is a great way to streamline the path from innovation to industry implementation.

References

[1] Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in US acute care hospitals. J Med Syst 2013;37(2):1-9.

[2] HIMSS Analytics United States EMR Adoption Model [Internet]. 2014. Chicago, Illinois: HIMSS. Available from: http://www.himssanalytics.org/emram/emram.aspx

[3] Gartner Hype Cycle [Internet]. 2014. Stamford, Connecticut: Gartner. Available from: http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

[4] Lee J, McCullough JS, Town RJ. The impact of health information technology on hospital productivity. Rand J Econ 2013;44(3):545-568.

[5] Murphy EV. Clinical Decision Support: Effectiveness in improving quality processes and clinical outcomes and factors that may influence success. Yale J Biol Med 2014;87(2):187.

[6] El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii40-ii51.

[7] Aarts J. A sociotechnical perspective of health information technology. Int J Med Inf 2013;82(12):1133-1135.

[8] Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13.

[9] Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review. Int J Med Inf 2013;82(5):e73-e86.

[10] Advocate Health Care business update and training for leaders. 2014. Oak Brook, Illinois: Advocate Health Care.

[11] Bamidis PD, Vivas AB, Styliadis C, Frantzidis C, Klados M, Schlee W, Siountas A, Papageorgiou SG. A review of physical and cognitive interventions in aging. Neurosci Biobehav 2014;44:206-220.

[12] Krive J, Shoolin JS, Zink SD. Effectiveness of evidence-based CPOE pneumonia order sets measured by health outcomes. J Public Health Inform. Forthcoming 2015.

Address for correspondence

Jacob Krive: krive@uic.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-414

Domain Analysis of Integrated Data to Reduce Cost Associated with Liver Disease

Tasneem Motiwala[a], Bobbie Kite[a], Kelly Regan[a], Gregg M. Gascon[b], and Philip R.O. Payne[a]

[a] *Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA*
[b] *The Ohio State University Health Plan, The Ohio State University Wexner Medical Center, Columbus, OH, USA*

Abstract

Liver cancer, the fifth most common cancer and second leading cause of cancer-related death among men worldwide, is plagued by not only lack of clinical research, but informatics tools for early detection. Consequently, it presents a major health and cost burden. Among the different types of liver cancer, hepatocellular carcinoma (HCC) is the most common and deadly form, arising from underlying liver disease. Current models for predicting risk of HCC and liver disease are limited to clinical data. A domain analysis of existing research related to screening for HCC and liver disease suggests that metabolic syndrome (MetS) may present oppportunites to detect early signs of liver disease. The purpose of this paper is to (i) provide a domain analysis of the relationship between HCC, liver disease, and metabolic syndrome, (ii) a review of the current disparate sources of data available for MetS diagnosis, and (iii) recommend informatics solutions for the diagnosis of MetS from available administrative (Biometrics, PHA, claims) and laboratory data, towards early prediction of liver disease. Our domain analysis and recommendations incorporate best practices to make meaningful use of available data with the goal of reducing cost associated with liver disease.

Keywords:
Liver cancer; liver disease; metabolic syndrome; economics.

Introduction

The availability of research, clinical and billing data provides an opportunity for the surveillance and management of chronic diseases and may lead to a reduction in associated costs. However, informatics challenges such as data accessibility, validation, integration, and lack of appropriate analytic tools inhibit progress. One such disease of interest that may benefit from domain analysis and data integration is liver disease.

Liver cancer is the fifth most common cancer and the second leading cause of cancer-related death in adult men worldwide [1]. The incidence of liver cancer in the United States has been rising slowly for the past few decades. According to estimates from the American Cancer Society, 33,190 new cases will be diagnosed and 23,000 people will die of this cancer in 2014 [2]. With an overall five-year survival rate of only ~16% [3], the mortality rate of liver cancer mimics its incidence rate. Unlike the decrease in mortality rates of other malignancies, mortality related to liver cancer has increased over the past two decades [4]. Late-stage detection, ineffective therapy, and other liver co-morbidities contribute to low survival and high mortality. Liver cancer is not only a public health burden, but also presents a significant fiscal burden on individuals and health plans with an estimated average first-year cost of more than $62,000 for treatment. The average first year post-transplant cost for liver transplant, currently the only potentially curative option for liver cancer and liver failure due to end-stage liver disease (see below), ranged from $158,274 - $248,358 for Medicare patients in 2008-09 [5].

Of the various types of liver cancer, hepatocellular carcinoma (HCC), involving uncontrolled proliferation of hepatocytes, is the most common, deadly, and costliest form of liver cancer accounting for ~80% of adult primary liver cancer [6]. The cost and associated poor prognosis of HCC make the cancer a prime candidate for early detection and intervention. HCC usually arises due to complications from underlying liver disease [7]. Therefore, using administrative and clinical measurement guidelines to characterize metabolic syndrome has the potential to offer a gateway to early liver disease detection and possible improvements in cost and patient outcomes related to HCC.

Liver cancer is one result of liver disease. The focus of this study is on a liver disease called cirrhosis, where cells become damaged and are replaced by scar tissue. Cirrhosis contributes as a major risk factor (condition that increases your risk of developing a disease) for HCC [4] in addition to demographic characteristics like age, gender and race/ethnicity. Current key causes of cirrhosis in the US include, but are not limited to: 1) *Alcohol abuse (alcoholic liver disease, ALD):* Chronic exposure to high levels of alcohol causes liver diseases ranging from simple hepatitis (inflammation) to steatohepatitis (fatty liver with inflammation) and consequently to cirrhosis; 2) *Chronic viral hepatitis:* Chronic infection of the liver with hepatitis B or hepatitis C virus causes chronic hepatitis that progresses to liver damage, which in turn leads to cirrhosis. In the US, 50% - 60% of HCC cases are infected with HCV, 10% - 15% are infected with HBV, less than 5% are infected with both, and 30% - 35% are not infected with any virus (cryptogenic HCC) [8]; and 3) *Non-alcoholic fatty liver disease (NAFLD):* NAFLD, like ALD, also refers to a spectrum of diseases that ranges from simple steatosis (fatty liver) to steatohepatitis (non-alcoholic steatohepatitis, NASH) and ultimately cirrhosis. However, unlike ALD, these diseases occur in the absence of exposure to excessive alcohol. The majority of cryptogenic HCC is thought to arise from NAFLD [9].

Understanding cirrhosis and its prevailing causes in the US is an important step in understanding and developing risk prediction models. Such models enable: (i) decisions regarding earlier or frequent screening; (ii) counseling for behavioral changes to decrease risk; (iii) designing chemoprevention and screening interventions; (iv) planning of

resource allocations for preventive measures; and (v) early disease management decisions. Thus, risk prediction models have the potential to aid individuals and their healthcare delivery in multiple ways.

During the past decade, several risk prediction models for HCC have been developed. A vast majority of these models are specifically applicable to those infected with either HBV or HCV. Different models of HCC prediction in HBV-infected individuals evaluate distinct combinations of factors like age, gender, HBV e antigen (HBeAg), HBV genotype, serum alanine transaminase (ALT), HBV DNA and HBV surface antigen (HBsAg), family history of HCC, alcohol consumption, albumin levels, bilirubin levels, and cirrhosis, core promoter mutation [10-13]. In contrast with the inclusion of virus-related measures for the prediction of HCC in HBV-infected individuals, factors associated with the prediction of HCC in HCV-infected individuals do not include virus serostatus or any other virus-related measures. Instead, unique combinations of factors like age, gender, platelet counts, albumin, aspartate transaminase (AST), liver stiffness (measured by transient elastography), and alpha-fetoprotein (AFP) concentration can predict risk of HCC in chronic HCV patients [14, 15]. A very recent study aimed at developing a risk prediction model for HCC in the general population has incorporated the etiology of cirrhosis as part of their model [4]. This model, called ADRESS-HCC, generates a risk score from the six factors: age, diabetes, race, etiology (autoimmune, alcohol/metabolic, viral), sex, and severity of liver dysfunction as measured by Child-Turcotte-Pugh (CTP) score. The resulting models are primarily restricted to individuals with HBV and HCV. However, about 65% of HBV- and 75% of HCV-infected individuals, respectively, are not aware of their infection status [16] precluding the assessment of their risk for liver cancer. Thus, risk of HCC in the general population needs to be better characterized. Further, these models have a tendency to largely neglect NAFLD, which is an emerging disease of importance.

The importance of NAFLD is evident from several facts [17]. NAFLD, which is associated with obesity and metabolic syndrome, is becoming epidemic with 31% of US population estimated to have this disease [18]. Approximately 1.5-2% of the US population has cirrhosis due to NAFLD. As a result, the burden of cirrhosis caused by NAFLD is twice as much as that caused by HCV. It is anticipated that within the next 5 years, NAFLD-related cirrhosis will surpass HCV-related cirrhosis as the leading indicator for liver transplant [17]. Liver cancer can develop in individuals with NAFLD in the absence of cirrhosis [19]. Presence of fatty liver increases risk of HCC associated with other liver diseases like chronic HCV and alcoholic liver injury. For example, risk of HCC in HCV infected individuals is 2-3-fold higher with coincident steatosis compared to those without steatosis [17]. It is unclear whether the process of neoplastic transformation is initiated after establishment of cirrhosis or during the early stages of liver disease, especially since steatosis (the earliest manifestation of NAFLD) has carcinogenic potential [7]. As a result, NAFLD could potentially be a more common predictor encompassing other liver diseases including ALD and HCV.

From the increasing prevalence of NAFLD and its association with hepatic and non-hepatic related mortality, it is apparent that screening and diagnosis of NAFLD is important for public health. Currently, liver biopsy is the gold standard for accurate diagnosis after suspected NAFLD to determine histological severity and prognosis of liver damage [20, 21]. Even though liver biopsies are the gold standard, they are performed with a particular set of risks [22]. While rare, complications of bleeding, infection and injury to nearby organs do exist with this invasive test and as such preclude its use as a screening tool in otherwise healthy individuals. Alternative methods of diagnosing NAFLD include ultrasonography, computed tomography and proton magnetic resonance spectroscopy (^{1}H-MRS). These techniques either lack sensitivity or are expensive and time-consuming [21]. To overcome these limitations, several prediction scores have been developed to encourage non-invasive assessment of NAFLD [21]. Current prediction methods and basis for their scores are:

1. *SteatoTest:* This test is a biochemical assessment of steatosis grade providing a numerical estimate ranging from 0.00 - 1.00 that corresponds to steatosis system of grades S0 -S4. It combines the six components of FibroTest/ActiTest [23] [alpha-2-macroglobulin (A2M), haptoglobin, apolipoprotein A1 (ApoA1), gamma-glutamyl transpeptidase (GGT), total bilirubin and ALT] with body mass index (BMI), serum cholesterol, triglycerides (TG) and glucose adjusted for age and gender.

2. *Fatty Liver Index (FLI):* The algorithm for FLI was developed based on the variables BMI, waist circumference, TG and GGT. The FLI ranges from 0 - 100. A FLI of less than 30 rules out fatty liver whereas FLI above 60 rules in fatty liver for further investigation by ultrasonography. It is thus a continuous measure that identifies the binary condition of fatty liver.

3. *NAFLD Liver Fat Score (LFS):* Among the many variables associated with increased liver fat content, metabolic syndrome (diagnosed according to the criteria defined by the International Diabetes Foundation [24]), increased fasting serum insulin, type 2 diabetes, fasting serum AST and decreased AST/ALT ratio were found to be independent predictors of NAFLD. These variables were used to calculate the NAFLD risk score. Since prediction of NAFLD is based on a specific cut score instead of a range of continuous values, this score does not reflect the severity of the liver disease.

4. *Lipid Accumulation Product (LAP):* This scoring mechanism, based on waist circumference and fasting TG, was originally used to detect prevalent cardiometabolic risk factors and diabetes [25, 26]. Since these two variables are common to FLI, LAP was tested for its ability to detect liver steatosis. Further, to avoid loss of information due to binary classification (as done in FLI), liver steatosis was evaluated as an ordinal 3-level variable. Nomograms depicting the probability of liver steatosis as a function of log-transformed LAP (lnLAP) indicated that the odds of more severe steatosis increased with increasing values of lnLAP.

5. *Hepatic Steatosis Index (HSI):* Lower BMIs and waist circumference in the Korean population compared to the Caucasian population precludes the application of FLI in this population. As a result, this novel index of NAFLD presence was devised by deriving an equation incorporating BMI, ALT/AST ratio and diabetes. A HSI of less than 30.0 rules out NAFLD whereas HSI of higher than 36.0 rules in NAFLD suggesting further screening. Due to the differences in BMI, waist circumference and susceptibility to insulin resistance between Korean and Caucasian populations, the cut-off values need to be reassessed for use of HSI in the US.

Independent development of these models limited generalizability of conclusions concerning superiority. Thus, a recent study was undertaken to compare the performance of these prediction scores on a single population in the US. For this purpose, data from the National Health and Nutrition Examination Survey (NHANES) conducted from 1988 to 1994 and subsequent follow-up data for mortality up to December 2006 were used [21]. According to this study, the performance of LFS for predicting NAFLD was superior to LAP, FLI or HSI. Further, LFS scores were also associated with outcomes. Higher LFS scores were associated with all-cause mortality. The high LFS group had a 31.25-fold and 30.3-fold increase in risk of liver mortality compared to the low and intermediate LFS groups, respectively. Cheung et al., 2014, suggest that the LFS model performs better than the others because it was derived using NAFLD cases diagnosed with ^{1}H-MRS, which is a very sensitive and quantitative imaging tool for identifying hepatic steatosis compared to ultrasonography that was used to derive the other three models. A difference noted in this method was the inclusion of metabolic syndrome in the algorithm, which is not a component of the other methods. It remains unknown whether or not the inclusion of metabolic syndrome creates a superior algorithm when compared to others.

Metabolic syndrome (MetS) is a cluster of risk factors that increases a person's risk for heart disease, stroke and NAFLD [46]. Three different organizations, the US National Cholesterol Education Program Adult Treatment Panel III (NCEP-ATP III), the World Health Organization (WHO) and the American Association of Clinical Endocrinologists (AACE), had initially recommended criteria for the diagnosis of MetS [27]. Subsequently, the International Diabetes Federation (IDF) and the American Heart Association/National Heart, Lung and Blood Institute (AHA/NHLBI) attempted to reconcile differences in clinical definitions of the recommended criteria but in turn outlined separate recommendations [28]. In general, the conditions that comprise metabolic risk factors include: 1) Obesity: large waistline (abdominal obesity) or high body mass index (BMI), 2) High TG level or on medicine for treatment of high TG, 3) Low high-density lipoprotein (HDL) or on treatment for low HDL, 4) High blood pressure or on treatment for high blood pressure, and 5) Insulin resistance: high fasting blood glucose, impaired glucose tolerance, type 2 diabetes or on treatment for high blood sugar.

Although the ideal thresholds and cut-off values differ between the different recommendations, the presence of three or more of these factors is required for the diagnosis of MetS [28]. According to the National Health and Examination Survey (NHANES) III (1988-1994) and NHANES 1999-2000, the age-adjusted prevalence of MetS was 24.1% and 27%, respectively [29]. It is now recognized that, in addition to MetS-associated risk of cardiovascular events and diabetes, MetS also impacts the liver in deleterious ways:

1. NAFLD is often associated with the individual components of MetS: ~90% of NAFLD patients have more than one feature of MetS and >33% have more than three criteria [30]. Further, a prospective cohort study conducted in Japan demonstrated that those with metabolic syndrome have a 4-11 times higher risk for future NAFLD [31] suggesting that NAFLD is a hepatic manifestation of MetS. Importantly, the presence of MetS is associated with progression of NAFLD to NASH and cirrhosis that may lead to HCC [32].
2. MetS is associated with faster progression of HCV-induced liver damage and decreased response to anti-viral therapy [32].
3. Obesity and overweight, components of MetS, are risk factors for the progression of ALD [33].
4. Risk of HCC is directly proportional to the number of MetS components present [34]. Importantly, HCC can develop even in the absence of cirrhosis when MetS and NAFLD are present [32, 35].

These relationships of MetS with liver disease and HCC strongly suggest that MetS can serve as an important predictor of liver disease (including NAFLD) and the probability of liver disease progressing to more complex conditions like cirrhosis and HCC.

Methods

We employed domain analysis of both research and disparate sources of data to develop recommendations for future directions towards cost reduction. We examine the informatics landscape from which these disparate data are generated to characterize MetS.

Results

We discovered that certain biomedical sources of data used to characterize MetS are not often used to screen for liver disease.

Current methods and uses of screening for MetS include:

1. *Data extraction from electronic medical records (EMRs):* A prospective cohort study developed an algorithm to identify patients at risk for cardiovascular health disease (CHD) or diabetes using data from EHRs [36, 37]. A validation study that compared electronic data extraction with manual chart review demonstrated substantial agreement between the two methods for identifying MetS. The prevalence of MetS determined by these methods was much higher than that determined from ICD-9 diagnosis code for MetS [38]. This is consistent with an earlier finding that metabolic syndrome is rarely coded in claims data using the designated ICD-9 code (dysmetabolic syndrome: 277.7) [39]. An internal analysis of claims data from The OSU Health Plan echoed this finding (where a 2013 incidence rate was calculated as 0.6 per 1,000).
2. *On-site health risk appraisal data of employed population:* A large study conducted on employees of a multistate financial services corporation used on-site health risk assessment that included a biometric screening by nurses and a questionnaire covering biological, psychological, and lifestyle health risks as well as presence of chronic diseases and health conditions and treatment for various medical conditions. Using this information, MetS (determined from biometrics data and treatment information) was found to be associated with poor perceived health, increased illness days and an increased trend of short-term disability incidence [40]. In a similar study conducted on employees of a manufacturing company, MetS was found to be associated with an increase in self-reported diseases (arthritis, chronic pain, diabetes, heartburn, heart disease and stroke) [41], and increasing healthcare and short-term disability cost [42].

Here, healthcare claims data was used to derive cost for the employees.

3. *Combination of multiple sources of information:* A retrospective analysis of a national EHR database containing clinical data, medication lists and prescriptions, and diagnoses or problems showed that clinical data identified more patients with cardiometabolic risk factors than either treatment- or diagnosis-based information [43]. Birnbaum et al. (2011) demonstrated that the use of claims or laboratory-based MetS patient identification alone underestimated MetS prevalence rates in comparison with an integrated data approach.

Recommendations: Screening

According to the claims data from The OSU Health Plan, HCC and end-stage liver diseases are among those conditions that consume the highest amounts of fiscal resources. A limitation of previous research is that despite the relationship between MetS and liver disease, no known studies have attempted to use MetS screening in a large population for understanding the risk of liver disease and HCC. Since risk factors for MetS are modifiable, identifying individuals with MetS and consequently, those at high risk for the liver diseases (including HCC), may facilitate intervention measures to alter these risks. Such early interventions at the MetS stage can prevent negative patient outcomes thus reducing the costs associated with treating the disease and continued monitoring for disease progression.

Next steps thus include screening and characterizing MetS for a population to encourage early interventions that can prevent or delay the onset of liver disease. Biometrics data for the past seven years (2008-2014), personal health assessments (PHAs) for the past two years (2013-2014), medication information from claims data for the past eight years (2007-2014), and laboratory values that are available through the Health Plan will be invaluable resources in the proposed screening of Health Plan members. Possible algorithms that might be developed using this data include:

1. *MetS risk prediction:* By restricting MetS characterization by yearly increments, we are able to determine incident rates based on clinically measured guidelines. Incident MetS cases can be determined from biometrics, PHA, laboratory test results, and medication data. Past data for matching individuals may be used to determine differentiating markers predicting risk by comparing data for those individuals without MetS. Those at risk for MetS can be referred to payer care-coordination programs for recommendations to control these risk factors.
2. *Liver disease risk prediction in presence of MetS:* MetS cases with or without subsequent ICD-9 diagnosis of liver disease can be identified using data from biometrics, PHA, laboratory tests and medications (for MetS diagnosis) as well as ICD-9 codes (for liver disease diagnosis). Differentiating features between these two groups can help to develop appropriate risk prediction algorithms. MetS has the potential to be measured as a continuous variable with specific cut-offs determining severity and risk of liver disease (similar to the algorithms developed for predicting NAFLD). Since there are different types of liver disease (viral, alcoholic and non-alcoholic), this analysis can be performed independently and collectively followed by testing to determine differences in performance of the prediction algorithm. Individuals identified to be at risk for liver disease can be followed-up with non-invasive tests (biochemical and image-based) to determine the presence and severity of the disease [44] in addition to referral to care coordination programs.
3. *HCC risk prediction in presence of MetS:* Similar to the above, retrospective data from identified MetS cases and/or those at risk for liver disease with subsequent ICD-9 codes for HCC can be compared with data for those without diagnosed HCC to determine differences in features. These features can be used to develop an algorithm that can accurately predict an individual's risk of HCC. Individuals identified at risk for HCC can be referred for liver biopsy for further confirmation of disease state.

The established chain of liver disease progression, presents a unique opportunity for informatics solutions to address results of this domain analysis and towards intervening in the resulting continuum of cost.

References

[1] Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69-90.

[2] What are the key statistics about liver cancer? [updated 02/10/2014; cited 2014 Dec 4]. Available from: http://www.cancer.org/cancer/livercancer/detailedguide/liver-cancer-what-is-key-statistics.

[3] Liver Cancer: Statistics: Cancer.Net; [cited 2014 Dec 4]. Available from: http://www.cancer.net/cancer-types/liver-cancer/statistics.

[4] Flemming JA, Yang JD, Vittinghoff E, et al. Risk prediction of hepatocellular carcinoma in patients with cirrhosis: The ADRESS-HCC risk model. *Cancer*. 2014;120(22):3485-93.

[5] Levi J, Pastorfield C, Segal LM, et al. HBV & HCV: America's Hidden Epidemics. Trust for America's Health, 2010.

[6] Liver Cancer: Overview: Cancer.Net; [cited 2014 Dec 4]. Available from: http://www.cancer.net/cancer-types/liver-cancer/overview.

[7] Kikuchi L, Oliveira CP, Carrilho FJ. Nonalcoholic fatty liver disease and hepatocellular carcinoma. *Biomed Res Int*. 2014;2014:106247.

[8] El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology*. 2012;142(6):1264-73 e1.

[9] Lee SS, Jeong SH, Byoun YS, et al. Clinical features and outcome of cryptogenic hepatocellular carcinoma compared to those of viral and alcoholic hepatocellular carcinoma. *BMC Cancer*. 2013;13:335.

[10] Lee MH, Yang HI, Liu J, et al. Prediction models of long-term cirrhosis and hepatocellular carcinoma risk in chronic hepatitis B patients: risk scores integrating host and virus profiles. *Hepatology*. 2013;58(2):546-54.

[11] Wong VW, Chan SL, Mo F, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *J Clin Oncol*. 2010;28(10):1660-5.

[12] Yang HI, Sherman M, Su J, et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *J Clin Oncol*. 2010;28(14):2437-44.

[13] Yuen MF, Tanaka Y, Fong DY, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. *J Hepatol*. 2009;50(1):80-8.

[14] Ikeda K, Arase Y, Saitoh S, et al. Prediction model of hepatocarcinogenesis for patients with hepatitis C virus-related cirrhosis. Validation with internal and external cohorts. *J Hepatol*. 2006;44(6):1089-97.

[15] Masuzaki R, Tateishi R, Yoshida H, et al. Risk assessment of hepatocellular carcinoma in chronic hepatitis C patients by transient elastography. *J Clin Gastroenterol*. 2008;42(7):839-43.

[16] Colvin HM, Mitchell AE. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C. Institute of Medicine, 2010.

[17] Michelotti GA, Machado MV, Diehl AM. NAFLD, NASH and liver cancer. *Nat Rev Gastroenterol Hepatol*. 2013;10(11):656-65.

[18] Lazo M, Clark JM. The epidemiology of nonalcoholic fatty liver disease: a global perspective. *Semin Liver Dis*. 2008;28(4):339-50.

[19] Yasui K, Hashimoto E, Tokushige K, et al. Clinical and pathological progression of non-alcoholic steatohepatitis to hepatocellular carcinoma. *Hepatol Res*. 2012;42(8):767-73.

[20] Adams LA, Angulo P. Recent concepts in non-alcoholic fatty liver disease. *Diabet Med*. 2005;22(9):1129-33.

[21] Cheung CL, Lam K, Wong I, et al. Non-invasive score identifies ultrasonography-diagnosed non-alcoholic fatty liver disease and predicts mortality in the USA. *BMC Med*. 2014;12(1):154.

[22] Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med*. 2001;344(7):495-500.

[23] Ratziu V, Massard J, Charlotte F, et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol*. 2006;6:6.

[24] Alberti KG, Zimmet P, Shaw J, et al. The metabolic syndrome--a new worldwide definition. *Lancet*. 2005;366(9491):1059-62.

[25] Kahn HS. The "lipid accumulation product" performs better than the body mass index for recognizing cardiovascular risk: a population-based comparison. *BMC Cardiovasc Disord*. 2005;5:26.

[26] Kahn HS. The lipid accumulation product is better than BMI for identifying diabetes: a population-based comparison. *Diabetes Care*. 2006;29(1):151-3.

[27] Grundy SM, Brewer HB, Jr., Cleeman JI, et al. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*. 2004;109(3):433-8.

[28] Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation*. 2009;120(16):1640-5.

[29] Ford ES, Giles WH, Mokdad AH. Increasing prevalence of the metabolic syndrome among u.s. Adults. *Diabetes Care*. 2004;27(10):2444-9.

[30] Marchesini G, Bugianesi E, Forlani G, et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. *Hepatology*. 2003;37(4):917-23.

[31] Hamaguchi M, Kojima T, Takeda N, et al. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. *Ann Intern Med*. 2005;143(10):722-8.

[32] Rosselli M, Lotersztajn S, Vizzutti F, et al. The metabolic syndrome and chronic liver disease. *Curr Pharm Des*. 2014;20(31):5010-24.

[33] Raynard B, Balian A, Fallik D, et al. Risk factors of fibrosis in alcohol-induced liver disease. *Hepatology*. 2002;35(3):635-8.

[34] Turati F, Talamini R, Pelucchi C, et al. Metabolic syndrome and hepatocellular carcinoma risk. *Br J Cancer*. 2013;108(1):222-8.

[35] Paradis V, Zalinski S, Chelbi E, et al. Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: a pathological analysis. *Hepatology*. 2009;49(3):851-9.

[36] Hivert MF, Dusseault-Belanger F, Cohen A, et al. Modified metabolic syndrome criteria for identification of patients at risk of developing diabetes and coronary heart diseases: longitudinal assessment via electronic health records. *Can J Cardiol*. 2012;28(6):744-9.

[37] Hivert MF, Grant RW, Shrader P, et al. Identifying primary care patients at risk for future diabetes and cardiovascular disease using electronic health records. *BMC Health Serv Res*. 2009;9:170.

[38] Hsiao P, Wood GC, Hartman T, et al. Using Electronic Data Extraction to Identify Subjects with Metabolic Syndrome: A Validation Using Manual Chart Review. *Clin Med Res*. 2011;9(3-4):168.

[39] Ford ES. Rarer than a blue moon: the use of a diagnostic code for the metabolic syndrome in the U.S. *Diabetes Care*. 2005;28(7):1808-9.

[40] Burton WN, Chen CY, Schultz AB, et al. The prevalence of metabolic syndrome in an employed population and the impact on health and productivity. *J Occup Environ Med*. 2008;50(10):1139-48.

[41] Schultz AB, Edington DW. Analysis of the association between metabolic syndrome and disease in a workplace population over time. *Value Health*. 2010;13(2):258-64.

[42] Schultz AB, Edington DW. The association between changes in metabolic syndrome and changes in cost in a workplace population. *J Occup Environ Med*. 2009;51(7):771-9.

[43] Brixner D, Said Q, Kirkness C, et al. Assessment of Cardiometabolic Risk Factors in a National Primary Care Electronic Health Record Database. *Value in Health*. 2007;10:S29-S36.

[44] Castera L, Vilgrain V, Angulo P. Noninvasive evaluation of NAFLD. *Nat Rev Gastroenterol Hepatol*. 2013;10(11):666-75.

Address for correspondence

Tasneem Motiwala, PhD, Depatment of Biomedical Informatics, The Ohio State University, tasneem.motiwala@osumc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-419

The Case for Conceptual and Computable Cross-Fertilization Between Audit and Feedback and Clinical Decision Support

Benjamin Brown[a,b], Niels Peek[a,b], Iain Buchan[a,b]

[a] *Health e-Research Centre, Farr Institute for Health Informatics Research, University of Manchester, UK*
[b] *Greater Manchester Primary Care Patient Safety Translational Research Centre, University of Manchester, UK*

Abstract

Many patients do not receive care consistent with best practice. Health informatics interventions often attempt to address this problem by comparing care provided to patients (e.g., from electronic health record data) to quality standards (e.g., described in clinical guidelines) and feeding this information back to clinicians. Traditionally these interventions are delivered at the patient-level as computerized clinical decision support (CDS) or at the population level as audit and feedback (A&F). Both CDS and A&F can improve care for patients but are variably effective; the challenge is to understand how the efficacy can be maximized. Although CDS and A&F are traditionally considered separate approaches, we argue that the systems share common mechanisms, and efficacy may be improved by cross-fertilizing relevant features and concepts. We draw on the Health Informatics and Implementation Science literature to argue that common mechanisms include functions typically associated with the other system, in addition to other features that may prove fruitful for further research.

Keywords:
Medical Audit; Quality Indicators; Healthcare Quality Improvement; Clinical Decision Support Systems.

Introduction

The prominence of evidence-based medicine has led to widespread acceptance of what consitutes good care. Implementing this evidence in clinical practice, however, is challenging – often referred to as the "second translational gap" Such problems lead to adverse outcomes for patients. For example, in the UK alone there are thought to be over 3,000 unnecessary strokes per year because patients with atrial fibrillation do not receive anticoagulant medication [1]. Barriers to implement evidence-based care may occur at different levels: individual patient-practitioner; provider team; provider organisation; and health system policy [2]. Health informatics interventions often attempt to address these barriers by highlighting to clinicians when patients may not receive care consistent with best practice evidence (e.g., in clinical guidelines) through analysis of patient data (e.g., from electronic health records [EHR]). When these interventions are delivered via computers during clinical encounters with patients, the interventions are called clinical decision support (CDS). When delivered outside clinical consultations and at the population-level, they are are typically described as audit and feedback (A&F). Systematic reviews of both types of intervention suggest they are moderately effective at ensuring patients receive improved care [3–5]. However, the reviews also suggest interventions are highly variable: sometimes the interventions work very well, and sometimes they do not [3–5]. The current challenge is therefore to understand how to maximize the efficacy. Although they are traditionally considered separate approaches, in this paper we suggest that computerized CDS and A&F share common mechanisms, and that their efficacy may consequently be improved by cross-fertilizing relevant features and concepts between each other.

To build our argument, we draw on literature from Health Informatics and from Implementation Science. Despite our focus on computer-based tools, we also draw on relevant evidence from non-computerized interventions. First we examine computerized CDS and A&F seperately: their functions, mechanisms and features associated with success. Next we consider their common aspects and provide a rationale for cross-fertilization. Finally, we provide examples of how this could be achieved both with functions typically associated with the other, in addition to other features that may improve their success. We end with a discussion on implications for future research and other tools that facilitate human interpretation of patient data.

Computerized Clinical Decision Support

CDS (both computerized and non-computerized) refers to a heterogeneous set of tools that can be broadly defined as "active knowledge systems which use two or more items of patient data to generate case-specific advice" [6]. This definition contrasts with passive knowledge systems, in which the user themselves must search the system [6]. Musen et al. classify CDS in three basic varieties [7]:

1. Patient-specific, situation-specific alerts, reminders, physician order sets, or other recommendations for direct action;
2. Information about the current clinical context to retrieve highly relevant online documents (e.g., *infobuttons*);
3. Organisation and presentation of information in a way that facilitates problem solving and decision making (e.g. graphical displays, documentation templates, structured reports).

We equate *computerized* CDS with the first variety. These are considered classic computerized CDS systems [7], and are arguably the most common [8,9]. Such systems provide custom-tailored assessments or advice based on patient-specific data (e.g., from EHRs or order entry systems) in consultation with a knowledge-base (usually best-practice evidence e.g. clinical guidelines), delivered via a computer to professionals at the point of care. Examples of these systems include [7,8]:

- Alerting clinicians if they are about to perform an action that may have adverse consequences (e.g., prescribing a macrolide antibiotic in a patient taking a statin);
- Reminding clinicians to perform a task (e.g., that a patient requires a cholesterol blood test);
- Suggesting management options for a particular patient based on their specific circumstances (e.g., suggesting changes in cholesterol-lowering treatment for a patient with high cholesterol).

Computerized CDS typically employs event-condition-action rules such as those in Arden syntax [7]. If a patient's data (the "event"; e.g., cholesterol level), meets criteria in accordance with the knowledge-base (the "condition"; e.g., >5mmol/L), then the CDS is trigerred (the "action"; e.g., suggestion of options for intensification of statin treatment). We exclude probabilistic CDS tools (e.g., Bayesian diagnostic systems) from our argument because they are usually not based on predefined clinical standards such as clinical practice gudelines.

We can surmise that computerized CDS attempts to improve compliance with best practice evidence by addressing barriers at the individual (patient-practitioner) level [2]. These include the health professional's lack of awareness or familiarity with the evidence, or their inertia of previous practice [10]. CDS works by making information available and visible to the health professional during the clinical encounter when action can be taken. However, CDS only works for patients that are encountered, and it is often ignored during time-pressured clinical encounters, or when a patient has an over-riding competing clinical priority. As a result, the efficacy of CDS is modest and highly variable. A recent Cochrane review of "on-screen, point of care computer reminders" demonstrated they improved processes of care by a median of 4.2% (interquartile range [IQR], 0.8% to 18.8%) [11]. Another review found only 58% of trials demonstrated an improvement in either processes of care or patient outcomes [3]. This review also demonstrated that CDS is more likely to be effective if it is delivered outside the EHR or order entry system, provides advice to patients as well as health professionals, and requires the user to articulate why they ignored a recommendation [3]. It has also been suggested that CDS may be variably effective because it does not target organisation-level barriers [12].

Audit & Feedback

A&F (both computerized and non-computerized) can be defined as "any summary of clinical performance of health care over a specified period of time" [13]. Other names for A&F include "clinical performance feedback," "performance measurement," and "quality measurement."

The audit part of A&F involves analysing data to produce a summary measure of clinical performance (interchangeably called a *quality indicator*, *performance measure*, or some combination of the two). Data may be obtained from medical records, computerized databases, or observations from patients [13]. Clinical performance is judged for a specified population according to accepted best practice (e.g., clinical guidelines). Quality indicators usually quantify clinical performance in a Donebedian classification as:

- Structural measures (e.g., number of nurses on a ward);
- Process measures (e.g.,proportion of eligible atrial fibrillation patients on anticoagulation); or
- Outcome measures (e.g., proportion of diabetic patients with good glycemic control (intermediate outcome) or number of myocardial infarctions per year per unit of population).

These indicators are generally calculated as proportions by comparing individual patients' data to the performance standard: the number of patients meeting the criteria form the numerator (e.g., number of children given a vaccination), and the total number eligible to meet the criteria form the denominator (e.g., number of children in the population eligible to receive the vaccine).

Feedback of audit results takes place after the clinical encounter, generally outside the clinical environment, and may target an individual, team or organisation. It may be delivered in a written, verbal or computerized format, and may include supporting materials, such as suggestions for improvement [5,13]. Feedback may also include *benchmarking* – comparison of recipients' performance with colleagues.

Traditionally, A&F was laboriously undertaken by the health professionals using paper medical records. However, widespread use of EHRs and web-based technologies means it is now much easier to undertake across multiple providers by external agencies such as governments or health service managers. As a result there is now an abundance of *computerized* A&F tools in healthcare systems around the world, variably termed *dashboards*, *benchmarking tools*, or *report cards*. Some are crude implementations of generic business intelligence software, others are more carefully developed for healthcare. These tools present information to health professionals (and often other audiences such as patients) via websites, computer applications or e-mail. Unlike non-computerized A&F, computerized A&F tools rarely make suggestions for improvement action to be taken by recipients.

Like CDS, A&F addresses barriers on the individual (patient-practitioner) level by making health professionals cognisant of their performance. As feedback is delivered outside the clinical encounter, it provides space and time for reflection and self-awareness, which ideally leads to behavior change. However, A&F also has the potential to address team and organisation-level barriers too, such as lack of resources and structural constraints [10] through the following ways:

- Feedback may be delivered to teams of clinicians and health care managers in addition to individual practitioners;
- Feedback provides recipients with a systematic and comprehensive view of entire patient populations served by a team or organisation, rather than only focusing on individual clinical encounters.

In addition to the space and time for reflection afforded by A&F, these factors encourage the formulation of service re-design plans for quality improvement. A limiting factor of A&F is that these plans must be formulated and undertaken, for which there must be sufficient time and resources. Consequently, like CDS, the efficacy of A&F is also modest and highly variable. The most recent Cochrane review of A&F demonstrated a median improvement in processes of care of 4.3% (IQR 0.5% to 16.0%) [5]. This review also demonstrated that A&F is more likely to be successful if the recipient is not performing well at baseline, and if feedback is

provided by a supervisor or senior colleague, regularly, in multiple formats with clear targets and an action plan [5]. It has also been suggested that A&F may be more effective if it includes individual patient-level data (in addition to population summaries) [14–16], individual clinician-level data (in addition to team- or organisation-level) [16,17], and if it is provided in a timely manner [17,18].

Rationale for Cross-Fertilization

Although computerized CDS and A&F have traditionally been considered separate approaches to quality improvement, we argue the above evidence (summarized in Table 1) suggests they are in fact highly related for the following reasons:

1. **They use the same substrates**: Both interventions use EHR data and compare the observed clinical workflow against a clinical standard (e.g. guidelines).
2. **They use analogous analytic methods:** The number of event-condition-action rules triggered in computerized CDS systems for a specific patient population are equivalent to the numerator value of quality indicators in a computerized A&F system. The total number of patients for which the event-condition-action rules could be applied is equivalent to the denominator.
3. **They use similar methods to effect behavior change:** Both feed back to recipients assessments of observed care versus a clinical standard.

As described above, it is established that CDS and A&F are moderately effective at improving patient care. The current research challenge is to therefore understand how their efficacies can be maximized [3,19]. We argue that given their similarities, there is a need to explore potential, systematic cross-fertilization and learning between them. In the following section we present evidence that cross-fertilization of their associated functionalities could increase their associated effectiveness. We therefore argue that this relationship should be exploited in systematic, computable ways.

Suggestions for Cross-Fertilization in Practice

Evidence-based Synergies of Typical Functions

Computerized CDS is more effective if delivered seperately from the EHR or ordering system [3], and may also be improved if it targets team and organisation-level barriers [12]. Both these are features typically associated with A&F (Table 1). For example, a computerized CDS system may only remind a user to order an annual thyroid function blood test for someone on long-term thyroxine if their EHR is opened during a clinical encounter. However, this only works for patients that are seen, and the reminder may be ignored in a time-pressured clinical environment. If this system was delivered outside the EHR, it would be possible to see all the patients who needed the blood test (like A&F), which may provide the time and space to formulate a plan to ensure they all had the blood test. Furthermore, providing this more population-based information may also help address some of the team and organisation-level barriers (like A&F), such as understanding that additional services may be needed to facilitate all the patients receiving the blood test e.g. providing additional phlebotomy clinics or new phlebotomy staff.

Computerized A&F is more effective if provided in a timely manner [17,18], with suggested action plans [5], individual patient-level data [14–16] and individual clinician-level feedback [16,17]. These are features typically associated with CDS (Table 1). For example, a computerized A&F system normally only highlights the proportion of hypertensive patients who have uncontrolled blood pressure. This requires searching for the patients and formulating action plans to improve performance, for which there may not be the resources or skills. Computerized A&F may therefore be improved if the summary also provides individual-level data on which patients have uncontrolled blood pressure, in addition to suggestions for improvement action (like CDS). This may include individual patient actions such as choices for medication optimization, but also team and organisation-level changes such as introducing a home blood pressure monitoring service or installation of a blood pressure machine in the clinic waiting room. Some computerized CDS systems already have similar functions to this termed “population registries”, however they are uncommon in practice [8] and only provide suggestions for individual patient actions. Furthermore, to address barriers at the individual patient-practitioner level, such as health professional’s lack of awareness or familiarity with clinical guidelines, requires knowledge of which clinicians need targeting. This is facilitated if feedback reports specify individual clinician performance in addition to their team or organisation (like CDS). This may also be more effective if provided close to the time of the clinical encounter (like CDS), when the experience is fresh in the clinician’s mind and the patient’s care is amenable to action, for example they have not left the hospital or moved address.

Conceptual Extension to Synergies of Other Functions

The evidence above suggests that functions typically associated with A&F are likely to improve the efficacy of computerized CDS, and vice versa, which re-inforces our argument that the two interventions are related. It also suggests that cross-fertilization of other features, for which there may not currently be supporting evidence, may be worth investigating. For example, computerized CDS may be more successful if it also provided population-level summaries like A&F *during* clinical encounters. One may hypothesize that when a computerized CDS system is triggered (e.g. for raised cholesterol), knowing the proportion of the eligible population for whom the alert would also fire (e.g. proprtion of patients with high cholesterol) will put the information in a broader context of clinical performance, re-inforcing the alert’s importance. This may more effectively motivate the recipient to take action, reducing alert fatigue [20].

This principle may also extend to other features that are not necessarily considered typical of either CDS or A&F, but that have been shown to improve their efficacy (Table 1). For example, computerized CDS is more effective if it provides advice for patients and if it requires a reason for over-riding its advice [3]. To our knowledge these features have not been investigated extensively in computerized A&F, though may improve its efficacy: providing advice to patients may improve adherence to medication and engagement with healthcare in the same way as CDS; and requiring users to justify why feedback is ignored may improve cognitive engagement with the quality improvement process. We suggest these and other areas in Table 1 may prove fruitful areas for further research.

We acknowledge there may be other features of computerized CDS and A&F interventions that are associated with improved efficacy, which have not been mentioned above. This may include features for which there are currently conflicting opinions or evidence, such as the use of benchmarking in A&F [15,19,21,22], or other features not yet

discovered. The application of behavioural change theory and use of qualitative evaluations in future research may help identify these additional features. This idea is beginning to gain traction in the A&F literature [19,22], but to our knowledge has not yet been mirrored in the CDS literature. We suggest that when new theories or features are discovered for either computerized CDS or A&F, they should be inclusively applied to both and their influence on efficacy assessed.

Table 1 – Comparison of computerized CDS and A&F

Feature	CDS	A&F
Data source	EHR	EHR
Analytic method	Event-condition-action rules	Quality indicators
Unit of analysis	Individual patient	Population
Delivery	During clinical encounter	Outside/after clinical encounter
Users	Individual clinicians	Individual clinicians, teams, organisations
Recommends improvement actions	Yes	No
Features associated with success	Delivered outside EHR Providing advice to patients Requiring over-ride reasons Targeting organisation-level barriers	Low baseline performance of recipient Feedback provided by supervisor/senior Regular feedback Multiple formats of feedback Clear targets and action plans fed back Individual level information (patient and clinician) Timely information

Discussion

We have studied the features of computerized CDS and A&F and argued that they they should be considered as related, rather than separate, approaches to healthcare quality improvement. In doing so, we have suggested their efficacy may be improved through the cross-fertilization of features typically associated with the other, and that future research should explore linking the two in a computable synergy.

Previous attempts to increase the understanding of (the effectiveness of) CDS and A&F have largely been limited to bottom-up aggregation of empirical evidence concerning a hetereogeneous set of intervention studies, with little success to date. Our approach advocates consideration of the mechanisms that underpin how they work and what they are trying to achieve, which we think will be more successful.

Our argument is not that computerized CDS and A&F should be used as 'multifacted' or 'co-' interventions; these terms suggest separate tools glued together. Our vision is that given their similarities these interventions should seamlessly incorporate successful features and other learning from each other. Interestingly, in support of this assertion, merely adding reminders to A&F has not been shown to impact efficacy [5], nor adding summary feedback to CDS [3], and there is little support from systematic reviews for multifaceted over single-component interventions in general [23]. Furthermore, we do not advocate that the empirical evidence relating to computerized CDS and A&F be simply combined to increase the power of meta-regressions in systematic reviews, as this belies the clear differences between them. Our argument is rather that a shared conceptualization can strengthen the generation and testing of specific hypotheses that draw on their shared mechanisms. We believe our understanding of both interventions can be advanced by considering the empirical evidence across them and borrowing evidential strength from adjacent areas where appropriate.

In addition to improving the efficacy of both computerized CDS and A&F, there are corollary benefits to their cross-fertilization. For example, the cross-fertilization encourages the development of common technical standards, which promises to save implementation time and effort [24]. Linking quality indicators to possible improvement actions also provides a more accurate measure of care quality, which is important if used for accountability purposes (e.g., performance-based payment). For example, it is more accurate to know the proportion of uncontrolled hypertensive patients prescribed suboptimal medication, rather than simply the proportion of uncontrolled hypertensives.

Although we have limited our discussion to CDS and A&F, our argument may well extend to a broader set of computerized interventions that facilitate clinician interpretation of patient data. Examples include range checks for laboratory test results, electronic checklists, and risk prediction tools. And although we excluded probablistic CDS systems from our discussion, they may also be relevant. For example there are arguments for the application of risk prediction tools in A&F [14] and a need for actionable suggestions (like CDS) in risk prediction [25]. Furthermore, there may be implications for non-computerized interventions too, as there are suggestions that A&F is most effective when there is external facilitation [26], which may be considered an 'educational outreach' feature [13].

A limitation of our argument is that the argument has occasionally drawn on evidence from non-computerized CDS and A&F, in addition to their computerized counterparts. This was necessary because some of the literature (particularly regarding A&F) does not distinguish between these two modes of delivery [5,6]. We believe our assertions transcend the distinction between computerized and non-computerized versions of these tools. Nevertheless, future research should empirically test whether our hypotheses regarding cross-fertilization hold in solely computerized settings. Our group has already started to do this by developing experimental computerized interventions [27,28].

Conclusion

We argue that computerized CDS and A&F systems are not separate but highly related approaches to quality improvement. We suggest that cross-fertilization of features and learning betweem them may improve their efficacy. We

have provided examples of how this may be achieved in computable ways, along with suggestions for future research.

Acknowledgments

Supported by the Wellcome Trust [104438/Z/14/Z] and UK Medical Research Council [MR/K006665/1].

References

[1] Cowan C, Healicon R, Robson I, et al. The use of anticoagulants in the management of atrial fibrillation among general practices in England. Heart 2013; 99(16):1166–72.

[2] Ferlie E, Shortel S. Improving the Quality of Health Care in the United Kingdom and the United States: A Framework for Change. Milbank Q. 2000;79(2):281–315.

[3] Roshanov PS, Fernandes N, Wilczynski JM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013;346:f657.

[4] Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. Effect of point-of-care computer reminders on physician behaviour: a systematic review. CMAJ 2010;182(5):E216–25.

[5] Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012;(6).

[6] Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: potential problems and solutions. Proc Annu Symp Comput Appl Med Care. 1991;3–7.

[7] Musen MA, Middleton B, Greenes RA. Clinical Decision-Support Systems. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 4th ed. London: Springer-Verlag; 2014. p. 643–74.

[8] Wright A, Sittig DF, Ash JS, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. J Am Med Inform Assoc 2011;18(3):232–42.

[9] Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012;157(1):29–43.

[10] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH. Why Don't Physicians Follow Clinical Practice Guidelines. JAMA. 1999; 282(15):1458-65.

[11] Shojania K, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database Syst Rev. 2009;(3).

[12] Goud R, van Engen-Verheul M, de Keizer NF, et al. The effect of computerized decision support on barriers to guideline implementation: a qualitative study in outpatient cardiac rehabilitation. Int J Med Inform 2010;79(6):430–7.

[13] Cochrane Effective Practice and Organisation of Care Group (EPOC). Data Collection Checklist. 2002.

[14] Ivers N, Barnsley J, Upshur R, et al. My approach to this job is one person at a time. Can Fam Physician. 2014;60:258–66.

[15] Søndergaard J, Andersen M, Kragstrup J, et al. Why has postal prescriber feedback no substantial impact on general practitioners' prescribing practice? A qualitative study. Eur J Clin Pharmacol 2002;58(2):133–6.

[16] Cresswell KM, Sadler S, Rodgers S, et al. An embedded longitudinal multi-faceted qualitative evaluation of a complex cluster randomized controlled trial aiming to reduce clinically important errors in medicines management in general practice. Trials 2012;13(1):78.

[17] Hysong SJ, Best RG, Pugh J. Audit and feedback and clinical practice guideline adherence: making feedback actionable. Implement Sci 2006;1:9.

[18] Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual Saf Heal Care 2004;13:26–32.

[19] Ivers NM, Sales A, Colquhoun H, et al. No more "business as usual" with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci 2014;9:14.

[20] Kesselheim AS, Cresswell K, Phansalkar S, et al. Clinical decision support systems could be modified to reduce "alert fatigue" while still minimizing the risk of litigation. Health Aff 2011;30(12):2310–7.

[21] Kiefe CI, Allison JJ, Williams OD, et al. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA 2001;285(22):2871–9.

[22] Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. Implement Sci 2012;7(1):114.

[23] Squires JE, Sullivan K, Eccles MP, et al. Are multifaceted interventions more effective than single-component interventions in changing health-care professionals' behaviours? An overview of systematic reviews. Implement Sci 2014;9(1):152.

[24] Kukhareva P, Kawamoto K, Shields DE, et al. Clinical Decision Support-based Quality Measurement (CDS-QM) Framework: Prototype Implementation, Evaluation, and Future Directions. AMIA Annu Symp Proc (In press).

[25] Lewis GH. "Impactibility Models": Identifying the Subgroup of High-Risk Patients Most Amenable to Hospital-Avoidance Programs. Milbank Q 2010;88(2):240–55.

[26] Cannell PJ. Evaluation of the end user (dentist) experience of undertaking clinical audit in a PCT-led NHS Modernisation Agency pilot scheme. Prim Dent Care 2009;16(4):168–76.

[27] Brown B, Williams R, Sperrin M, et al. Making Audit Actionable: An Example Algorithm for Blood Pressure Management in Chronic Kidney Disease. AMIA Annu Symp Proc (In press).

[28] Van Engen-Verheul M, de Keizer N, van der Veer S, et al. Evaluating the effect of a web-based quality improvement system with feedback and outreach visits on guideline concordance in the field of cardiac rehabilitation: Rationale and study protocol. Implement Sci (In press).

Address for correspondence

Benjamin Brown, Health e-Research Centre, Farr Institute for Health Informatics Research, The University of Manchester, Vaughan House, Portsmouth Street, Manchester, M13 9PL, UK. benjamin.brown@manchester.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-424

Inside the Black Box of Audit and Feedback: a Laboratory Study to Explore Determinants of Improvement Target Selection by Healthcare Professionals in Cardiac Rehabilitation

Wouter T. Gude[a], Sabine N. van der Veer[b], Mariëtte M. van Engen-Verheul[a], Nicolette F. de Keizer[a], Niels Peek[b,c]

[a]*Department of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands*
[b]*Health e-Research Centre, Institute of Population Health, University of Manchester, United Kingdom*
[c]*NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, University of Manchester, United Kingdom*

Abstract

Audit and feedback (A&F) is widely used to aid healthcare professionals in improving clinical performance, but there is little understanding of the underlying mechanism that determines its effectiveness. The aim of this paper is to investigate the process by which healthcare professionals select indicators as improvement targets based on A&F. We performed a laboratory study among 41 healthcare professionals in the context of a web-based A&F intervention designed to improve the quality of cardiac rehabilitation care in the Netherlands. Feedback was provided on eighteen quality indicators, including a score and a colour (representing a recommendation for selection (red and yellow) or non-selection (green)). Indicators with more room for improvement were more likely to be selected, although this varied substantially between participants. In more than a quarter of the cases, participants did not select indicators with obvious room for improvement (yellow or red colour), or selected indicators without apparent room for improvement (green colour). We conclude that personal preferences and beliefs concerning quality and performance targets may dilute the efficiency of A&F.

Keywords:

Feedback, Psychological; Medical Audit/standards; Heart Diseases/rehabilitation.

Introduction

Healthcare organisations increasingly adopt audit and feedback (A&F) strategies to monitor and improve their quality of care. A&F provides healthcare professionals with an objective summary of their clinical performance over a specified period of time.[1] Clinical performance is typically measured by a set of quality indicators derived from clinical guidelines or expert opinion–each indicator representing a quality aspect of care (e.g., proportion of patients receiving a treatment according to guideline recommendations, or mortality rates). Despite the widespread use of A&F and the inherent efforts and costs put into their development and application, A&F interventions show variable effectiveness on improving quality of care. A recent Cochrane review of 140 randomised trials of A&F interventions reported a median 4.3% absolute improvement (interquartile range 0.5% to 16%) in quality of care, with a quarter of the studies showing a strong positive effect, but with another quarter showing a negative or null effect.[1]

Previous studies have attributed much of the observed variability in effect of A&F interventions to feedback design characteristics and contextual factors. They suggested A&F to be most effective if provided by a supervisor or colleague, more than once, both verbally and in writing, if baseline performance is low, and if it includes explicit targets and an action plan.[1–4] Other suggested effect modifiers are the perceived quality of the data underlying the feedback, motivation and interest of the recipient, organisational support for quality improvement (QI),[5] and how performance targets or benchmarks are derived.[6] However, whereas these studies contribute to our knowledge about what factors may influence the impact of feedback, they do not help us understand the underlying mechanisms of how A&F interventions affect the quality of healthcare.[7] Little progress has been made on this matter because most randomised controlled trials (RCTs) on A&F interventions did not explicitly build on previous research or extant theory.[8,9] Instead, they treated the intervention as a 'black box', focusing solely on outcomes (i.e., a change in quality of care) while ignoring the mechanism inside.

Control Theory is an increasingly recognised theory in A&F literature that offers an explanation of how this feedback mechanism works. It starts from the assumption that healthcare professionals are prompted to improve practice when observing a discrepancy between their own performance and a certain target (e.g., peer performance).[10] Feedback reports will continue to prompt improvement actions until the discrepancy has been solved, i.e., denoting target achievement. However, if professionals disagree with the target or observe a discrepancy that is too great, or professionals lack skills, motivation or resources for action, they may decide against selecting particular indicators as targets for improvement.[10,11] Herewith, Control Theory reveals two steps in the feedback mechanism essential to improving upon indicators: (i) healthcare professionals must recognise achievable room for improvement on indicators before selecting them as QI targets, and (ii) they must formulate and perform effective improvement actions. Understanding how these steps work is imperative for designing successful A&F interventions [12]; in this paper we investigate the first step.

The aims of this study were thus to determine (i) how healthcare professionals select quality indicators as targets for improvement based on reports of performance feedback, and (ii) what are reasons to disregard feedback's recommendations for selecting indicators. The study was conducted in a laboratory setting among cardiac rehabilitation professionals.

Methods

Study context: the audit and feedback intervention

The context of our study was the CARDSS Online A&F intervention, which was developed to improve concordance with guidelines and facilitate organisational change in cardiac rehabilitation (CR) in the Netherlands.[13] CR is a multidisciplinary therapy designed to support coronary heart disease patients recovering from a cardiac incident, and aims to improve their overall physical and psychosocial condition.[14] It is offered by multidisciplinary teams which generally include cardiologists, physical therapists, nurses, psychologists, dieticians, and social workers. A cluster-randomised trial among eighteen CR clinics in 2013 and 2014 evaluated its effectiveness [15]; the results will be published elsewhere. In the trial, local QI teams – consisting of CR professionals from various disciplines – received four quarterly feedback reports and educational outreach visits. During these four visits, the teams used CARDSS Online to select indicators as targets for improvement and to establish a QI plan consisting of concrete improvement actions and goals that were linked to the selected indicators. In follow-up visits, the teams assessed the new feedback and updated the QI plan accordingly. An average QI team consisted of 7 members (range 3 to 13). They were always part of a clinic's larger CR team, which consisted of a total of 6 to 16 professionals.

The feedback reports generated by CARDSS Online consisted of performance scores on eighteen structure, process, and outcome indicators for CR, combined with coloured icons to represent the objective room for improvement and a recommendation for selecting (red and yellow) or not selecting (green) an indicator as a target for improvement (see example feedback report in Figure 1). In the current study we focused on process and outcome indicators that reflected a dichotomous measure at the patient level (e.g., percentage of patients who quit smoking after CR). The indicator set was developed in close collaboration with CR professionals using a modified RAND method; the complete set is available in reference.[16] To determine performance scores for the indicators, all data were automatically derived from an electronic health record at the point of care. The coloured icons were assigned to an indicator based on the clinic's performance score and the average score across all clinics. Green was assigned if the clinic's score was ≥ 66%, except if the average score was ≥ 66%. Then the threshold for assigning a green colour was the average score minus 10%. If green was not assigned, yellow was assigned if the clinic's score was ≥ 33% or less than 10% below average. All other scores resulted in assignment of a red colour. In the case that a CR clinic submitted insufficent data to determine the score on a specific indicator, the feedback report lacked a score and showed a grey colour.

Study design and participants

We conducted our laboratory study among individual CR professionals from the eighteen CR clinics that were enrolled in the CARDSS Online trial. The 132 members of all local QI teams in the trial were invited by e-mail with up to two reminders (after 2.5 and 5 weeks).

All invited CR professionals received a personal account for CARDSS Online. For each professional, we randomly selected two real performance feedback reports from a report database containing all 50 reports that were previously generated for CR clinics participating in the CARDSS Online trial. We presented them as if they concerned reports of a 'virtual' CR clinic. Respondents were asked to select the quality indicators that they thought should be improvement targets in the QI plan of the virtual clinic. We did not set a limit on the number of selectable indicators, but encouraged respondents to compose a feasible QI plan that was to be evaluated after three months.

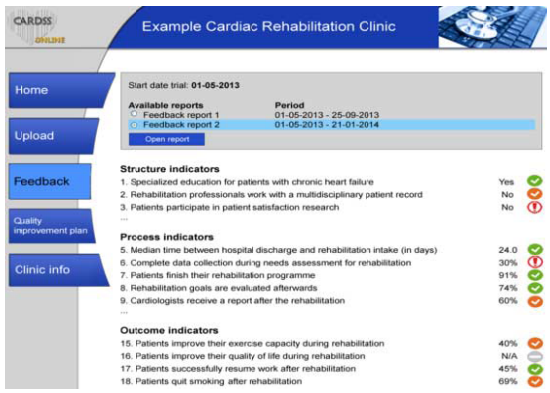


Figure 1 – Example feedback report in CARDSS Online

Data collection

Per quality indicator in each feedback report we collected the type of indicator (process or outcome), reported score (0 to 100%), reported colour (red, yellow, green, or grey), and whether the respondent selected the indicator as target for improvement (yes or no). To determine reasons for disregarding feedback's recommendations for (not) selecting indicators as improvement targets, we asked respondents to explain their choice if they chose not to select an indicator with a red or yellow colour by selecting one of the following statements: "not a relevant aspect of quality of CR", "improvement is not feasible", or "indicator score is high enough".[11] Similarly, we asked them to explain their choice if they did select an indicator with a green colour, using one of the following statements: "essential aspect of quality of CR", "easy to improve", and "indicator score is too low". If no statement applied, respondents could explain their choice in a text field.

At the level of the CR professional, we collected gender, age, years of clinical experience, discipline, whether they occupied a coordinating function in CR care, percentage of their time they spent on direct patient care, and how many educational outreach visits they attended in the CARDSS Online trial.

Statistical analysis

We performed three separate mixed-effects multivariate logistic regression analyses to determine which factors influence the selection of quality indicators by CR professionals. First, to test whether increased room for improvement was associated with higher chance of indicator selection (as posed by Control Theory), we used 'indicator type', 'reported score', and all collected professional-level factors as fixed covariates. Second, to determine whether the reported colour had additional or stronger influence, we performed another analysis by extending the model with the covariate 'reported colour'. Indicator values for which no score was available due to insufficient data (grey colour in feedback report) were left out of both analyses. Finally, to determine the impact of lacking feedback (grey colour) on indicator selection, we performed a third, similar analysis on the complete set of indicator values while leaving out the covariate 'reported score'. To account for clustering effects due to correlations between observations on the same quality

indicator (some indicators may be generally preferred to others), the same CR professional (some professionals may select more indicators than others), and the same clinic (professionals from the same clinic may have similar preferences), we added random intercepts for (i) indicator, (ii) respondent, and (iii) CR clinic to our regression models.

Following Nakagawa et al.'s [17] recommendations for reporting explained variance in mixed-effects models, we calculated two values of R^2: marginal and conditional R^2. Marginal R^2 describes the proportion of variance explained by the fixed effects alone, whereas the conditional R^2 describes the proportion of variance explained by both the fixed and random effects. For all fixed effects, we calculated the odds ratios (ORs) and 95% confidence intervals (CIs). To quantify how much of the explained variance could be attributed to each random effect, we substracted for each random effect the conditional R^2 of a model with only that single random effect from the marginal R^2 of the full model. All analyses were performed using R version 3.1.1 (R Foundation for Statistical Computing; Vienna, Austria).

Results

Fourty-one individual CR professionals (response rate, 31%) accepted our invitation; Table 1 summarises their characteristics. The majority of respondents were CR nurses and physiotherapists, had a coordinating function, and reported to spend more than half of their time on direct patient care. On average they attended approximately two educational outreach visits in the CARDSS Online trial. Two of the eighteen CR clinics from the clinical study were not represented in the study sample.

Table 1 – Participant characteristics (n=41)

Characteristic	Value*
Male gender	19 (46)
Mean age in years (SD)	45.6 (10.7)
Mean clinical experience in years (SD)	23.3 (10.5)
Mean visits attended (min-max)	1.83 (0-4)
Discipline	
CR nurse	18 (44)
Physiotherapist	13 (32)
Psychologist	3 (7)
Social worker	2 (5)
Dietician	1 (2)
Sports physician	1 (2)
Cardiologist	2 (5)
Manager / head of CR department	2 (5)
Coordinating function	22 (54)
Time spent on direct patient care	
Less than 25%	6 (15)
25% to 50%	9 (22)
50% to 75%	14 (34)
More than 75%	13 (32)
Mean no. of participants per clinic (SD)	2.7 (1.9)

* Values are numbers (percentages) unless indicated otherwise

In total, participants evaluated 82 feedback reports (two each). These evaluations concerned 40 unique reports (out of the 50 reports in the database), which were evaluated by one to four participants. The 82 reports contained a total of 767 patient-level dichotomous indicator values. 437 indicator values (57%) were assigned a grey colour due to a lack of data in calculating a score. Of the remaining 330 values, 287 (87%) were process indicators and 43 (13%) concerned outcome indicators. On average, participants selected 75% (range 52% to 100%) of all patient-level dichotomous indicators presented within the two feedback reports as targets for improvement. Participants most frequently selected the following indicators: proportion of patients who finish their relaxation and stress management training (process indicator; n=35, 100%), proportion of patients whose cardiovascular risk factors are evaluated after rehabilitation (process indicator; n=41, 100%), and proportion of patients who successfully resume work after rehabilitation (outcome indicator; n=39, 95%). These were also the indicators that most often showed a grey colour. Least frequently selected indicators were process indicators: proportions of patients for whom all data are collected during the needs assessment concerning lifestyle factors (n=5, 19%), psychological functioning (n=7, 26%), and proportion of patients that were offered a complete and tailored CR programme (n=30, 37%). These were also the indicators that most often showed a green colour.

Determinants of selecting quality indicators as targets for improvement

In the first analysis (Table 2), we found the reported performance score to be a significant determinant of selecting indicators as targets for improvement (OR 1.50, 95% CI 1.28 to 1.76; per 10% decrease in score). We found no association between the indicator type nor any of the professional-level factors and the selection of an indicator as target for improvement. The model explained 50.6% of the total variance (conditional R^2), where 20.8% was explained by the fixed effects alone (marginal R^2). Selection of indicators depended significantly on the participant, with its random effect accounting for 17.7% of the variance explained. Random effects of the CR clinic and quality indicator showed no significant influence on the selection of indicators.

In the second analysis, we found a significant influence of the reported colour on indicator selection, whereas the influence of the reported score was no longer significant. Red and yellow indicators were more often selected compared to green indicators (red: OR 10.42, 95% CI 1.28 to 84.85; yellow: OR 17.44, 95% CI 5.92 to 51.32). The total explained variance increased to 58.0% (conditional R^2), where 31.9% was explained by the fixed effects alone (marginal R^2).

The third analysis revealed that grey indicators were selected 100% of the time.

Reasons to disregard feedback's recommendations for selecting quality indicators

For the 330 of 767 quality indicator values with a reported score and colour, respondents deviated from the feedback's recommendation in 88 (27%) cases; they either did not select indicators showing a red or yellow colour (n=55, 34%), or they selected indicators showing a green colour (n=33, 19%).

Respondents' main reasons for not selecting red or yellow indicator values were that they reckoned the reported score to be high enough (n=16, 29%; fourteen yellow and two red indicators), and that improving the indicator was not feasible (n=14, 25%; in particular outcome indicator 'proportion of patients that quit smoking' [n=4] and process indicator 'proportion of patients whose rehabilitation goals are evaluated after CR' [n=3]). Further, they thought that improving the indicator lacked priority (n=11, 21%), or that the indicator did not represent a relevant quality aspect of CR (n=8, 15%).

The main reason for respondents' selecting green indicator values was that they considered the indicator an essential aspect of the quality of CR care, and should thus belong in every quality improvement plan (n=27, 82%; e.g., process indicator 'proportion of patients that were offered a complete

and tailored CR programme' [n=8]). All other reasons were selected in less than 10% of cases.

causes, making it more challenging to formulate effective improvement actions.

Table 2 – Determinants of selecting indicators as targets for improvement by CR professionals

Variable		p-value
Fixed effects	*Odds ratio (95% CI)*	
Quality indicator level		
Outcome vs. process indicator	1.01 (0.30, 3.39)	0.955
Reported score (per 10% decrease)	1.50 (1.28, 1.76)	<0.001
CR professional level		
Male vs. female gender	2.16 (0.42, 11.09)	0.363
Years of clinical experience (per 10 years)	1.10 (0.60, 1.99)	0.764
Spends more than half of the time on direct patient care	1.64 (0.53, 5.14)	0.393
Number of educational outreach visits attended in trial	1.03 (0.64, 1.65)	0.903
Discipline		
Physiotherapist vs. CR nurse	1.03 (0.13, 8.35)	0.969
Other vs. CR nurse	0.86 (0.18, 4.12)	0.848
Coordinating function	1.27 (0.45, 3.63)	0.650
Random effects	*Explained variance*	
CR clinic	4.4%	0.501
CR professional	17.7%	0.002
Quality indicator	7.7%	0.071

Discussion

We explored CR professionals' decision-making behaviour with regard to selecting quality indicators as improvement targets when confronted with performance feedback. Based on Control Theory, this is the first and essential step in the feedback mechanism to enable effective A&F. Professionals were more likely to select an indicator if there was more room for improvement, and they were more strongly influenced by the reported colour (red, yellow, green) than by the reported performance score. Still, they ignored more than a quarter of the feedback's recommendations for (not) selecting indicators based on the reported colours. Indicators for which insufficient data were available to calculate a score (indicators with a grey colour) were always selected. In addition, the number of selected indicators varied strongly between individual participants.

Relation to other studies

Although we are the first to have empirically investigated step 1 in the underlying feedback mechanism, earlier studies aimed at identifying design characteristics and contextual factors that influence the outcome of the mechanism (i.e., change in quality of care). A Cochrane review [1] concluded that low performance scores increased A&F effectiveness. The authors attributed this to greater intention to take action, or absence of ceiling effects [12]; our findings seem to support the first. Two reviews [2,3] found that feedback is more effective when accompanied by explicit performance targets. Supporting those results, our study revealed that CR professionals act more strongly upon colours, which served as a graphical representation of whether a target was achieved, than upon their absolute performance scores. Another review [5] suggested that A&F is more likely to improve process indicators compared to outcome indicators. However, our study found no effect of the type of indicator on the selection process. This implies that feedback's limited effect on outcome indicators may be explained by factors related to subsequent steps in the feedback mechanism outside our scope. For example, suboptimal scores on outcome indicators may require further drilling down to understand the underlying

Meaning of the study

A&F interventions often achieve only limited improvement of quality of care, for which current scientific knowledge fails to offer a satisfactory explanation. Our findings suggest that one of the reasons could be that feedback may not always convince healthcare professionals that certain aspects of their practice require improvement. In one third of the cases where quality indicators showed room for improvement (by showing a red or yellow colour), professionals did not select them as targets for the QI plan; they either disagreed with the performance target (they thought their score was high enough, thus not acknowledging room for improvement), deemed improvement unfeasible or lacking priority, or did not consider the indicator an essential aspect of care quality. Moreover, as this study was conducted in a laboratory setting amongst individual professionals, we expect that contextual factors in clinical practice (e.g., organisational constraints within a clinic, or social context within care teams) may further distort the process of selecting improvement targets based on performance feedback.

Nevertheless, our results show that using colours to represent the extent to which performance targets are achieved can be a useful means to guide healthcare professionals in selecting those indicators that require most improvement. This was illustrated by the fact that in our study red and yellow indicator values were much more often selected than green ones. Professionals also always selected indicators for which no feedback was available due to insufficient data. A possible explanation is that professionals seek to optimise data collection in order to draw a more complete picture of their clinical performance.

The number of selected indicators differed substantially between CR professionals. Our analysis did not reveal any systematic professional-level factors that may offer an explanation for this phenomenon. Instead, part of the explanation may come from individual professionals having personal preferences and beliefs concerning which indicators constitute care quality, and about what are achievable performance targets. Hence, accompanying feedback with explicit performance targets or colours may not always succeed in influencing the process of selecting improvement targets.

Limitations

Our study found reported colour to be a powerful tool to persuade healthcare professionals into selecting an indicator as target for improvement. Although there are various ways to derive colour thresholds for reported performance scores (such as based on peer performance, expert knowledge or clinical practice guidelines), the A&F design in our study used a relatively simple colour assignment scheme. This might partly explain why professionals in our study ignored a third of the feedback's recommendations for selecting indicators, in particular for the reasons "indicator score is high enough" and "improvement is not feasible" (both stated in almost one third of cases). In case our A&F design would have used a more sophisticated method for assigning colours, this proportion may have turned out smaller.

Furthermore, due to practical reasons, we only analysed the selection of indicators that reflected a dichotomous measure at the patient level; thus leaving out continuous measures and clinic-level dichotomous measures. Therefore we do not know whether the type of measure plays a role in professionals' selection of indicators as targets for improvement.

Future work

Currently we are conducting a clinical study to investigate how teams of CR professionals participating in the CARDSS Online trial select indicators as targets for improvement in clinical practice. Comparing the results of the clinical study to those in our laboratory study will provide insight as to the impact of real-world barriers on selecting indicators. Further research may also focus on investigating the second step in the feedback mechanism to explore determinants of actual change in performance.

Conclusion

CR professionals presented with indicator-based performance feedback were more likely to select indicators if there was more room for improvement, and they were more strongly influenced by the reported colour (red, yellow, green) than by reported performance score. Still, in more than a quarter of the cases they ignored the feedback's recommendations for (not) selecting indicators based on the reported colours. We conclude that personal preferences and beliefs concerning quality and performance targets may dilute the efficiency of performance feedback. A follow-up study will be conducted to investigate the same step in clinical practice among CR teams.

References

[1] Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012;6:CD000259. doi:10.1002/14651858.CD000259.pub3.

[2] Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. Med Care 2009;47:356–63. doi:10.1097/MLR.0b013e3181893f6b.

[3] Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: The example of audit and feedback. Soc Sci Med 2010;70:1618–25. doi:10.1016/j.socscimed.2010.01.039.

[4] Foy R, MacLennan G, Grimshaw J, Penney G, Campbell M, Grol R. Attributes of clinical recommendations that influence change in practice following audit and feedback. J Clin Epidemiol 2002;55:717–22.

[5] Van der Veer SN, de Keizer NF, Ravelli ACJ, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. Int J Med Inform 2010;79:305–23. doi:10.1016/j.ijmedinf.2010.01.011.

[6] Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA 2001;285:2871–9.

[7] Foy R, Eccles MP, Jamtvedt G, Young J, Grimshaw JM, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. BMC Health Serv Res 2005;5:50. doi:10.1186/1472-6963-5-50.

[8] Sales A, Smith J, Curran G, Kochevar L. Models, strategies, and tools: Theory in implementing evidence-based findings into health care practice. J Gen Intern Med 2006;21. doi:10.1111/j.1525-1497.2006.00362.x.

[9] Colquhoun HL, Brehaut JC, Sales A, Ivers N, Grimshaw J, Michie S, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. Implement Sci 2013;8:66. doi:10.1186/1748-5908-8-66.

[10] Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982;92:111–35. doi:10.1037/0033-2909.92.1.111.

[11] Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996;119:254–84. doi:10.1037/0033-2909.119.2.254.

[12] Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more "business as usual" with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci 2014;9:14. doi:10.1186/1748-5908-9-14.

[13] Van Engen-Verheul MM, Van Der Veer SN, De Keizer NF, Tjon Sjoe Sjoe W, Van Der Zwan EPA, Peek N. A Web-based system to facilitate local, systematic quality improvement by multidisciplinary care teams: Development and first experiences of CARDSS online. Stud. Health Technol. Inform., vol. 192, 2013, p. 248–52. doi:10.3233/978-1-61499-289-9-248.

[14] Fleming DM, Cross KW, Barley MA. Recent changes in the prevalence of diseases presenting for health care. Br J Gen Pract 2005;55:589–95.

[15] Dutch Trial Register. Implementation of the Cardiac Rehabilitation (CR) guidelines by a multifaceted decision support, feedback and outreach visit intervention: A cluster randomized trial. 2012. http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3251 (accessed December 10, 2014).

[16] Van Engen-Verheul M, Kemps H, Kraaijenhagen R, De Keizer N, Peek N. Modified rand method to derive quality indicators: A case study in cardiac rehabilitation. Stud. Health Technol. Inform., vol. 169, 2011, p. 88–92. doi:10.3233/978-1-60750-806-9-88.

[17] Nakagawa S, Schielzeth H. A general and simple method for obtaining R 2 from generalized linear mixed-effects models. Methods Ecol Evol 2013;4:133–42. doi:10.1111/j.2041-210x.2012.00261.x.

Address for correspondence

Wouter Gude, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, P.O. Box 22700, 1100 DE Amsterdam, The Netherlands. E-mail: W.T.Gude@amc.uva.nl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-429

Using Patient Journey Modelling to Visualise the Impact of Policy Change on Patient Flow in Community Care

Rosemary Sayvong[a], Dr Joanne Curry[ab]

[a]*University of Western Sydney, New South Wales, Australia*
[b]*CSC, Australia*

Abstract

Health policy plays a crucial role in community care, particularly within care programs such as ComPacks. ComPacks is a short-term care program administered by New South Wales (NSW) department of health which runs for up to 6-weeks and its goal is to prevent or minimise hospital readmission. Compliance to the ComPacks health policies is required in order to gain financial support from overnment bodies, however when the Government makes changes to service policies, this may potentially cause ripple effects to the workflow of the service and increase pressures on care providers, which in turn may affect the patient. Utilising a multi-layer visualisation tool can help identify whether changes made to policy are impacting patient flow in a positive or negative way. This research study investigates the use of an emerging patient journey modelling technique to better understand service design in a community care setting, whilst also determining the impact of State-level policy interventions.

Keywords:

Patient Journey Modelling; Essomenic; Health Policy; Community Care; ComPacks.

Introduction

The increasing prevalence of hospital readmissions and the subsequent impact on acute care services cannot be easily ignored. Easing the burden on acute care services can be managed through efficient and effective service delivery of community care services [1, 2]. *C*ommunity care programs have been initiated by Local and State government as a means to addressing community health issues. Such examples can be exemplified through programs such as Home Aged Community Care (HACC), Hospital in the Home (HITH), and Community Packages (ComPacks) in which patient care is transitioned from hospital to home and assistive services are provided to patients, post-discharge [1, 3, 4]. These programs in particular are driven by a focus on reducing the probability of readmissions due to post-discharge complications, thus reducing the pressures on hospital services. In understanding that state and local governments administer most of these care programs and services, policies and guidelines are developed and implemented as a directive measure in achieving an effective service delivery for care. These documents dictate service delivery processes and also indicate performance areas that are to be measured [5]. When changes are made to policy, it is difficult to examine whether the impacts of these changes are affecting service workflow, patient flow and patient outcomes without understanding the concepts of policy analysis. The following study examines the use of a healthcare-oriented service redesign technique as a tool to visualise the impact of policy change on service workflow, patient flow, and outcomes within community care.

Health Policy Development

Health policy is an essential component of the delivery of care services. It establishes a plan describing the courses of action or intended actions to be taken by institutions or organisations that have an impact on health [6–8]. Health policy is developed based on informed values and reasoning [5]. It is formulated, proposed, and evaluated using evidence-based methods to ensure successful implementation. Therefore, policy changes must undergo comprehensive analysis and evaluation. Analysis and evaluation have the potential to (1) increase policy impact and (2) provide evidence for future policy amendments [9, 10]. Based on literature surrounding existing theoretical frameworks for policy analysis, it is generally agreed that the policy development environment comprises several interconnecting elements; policy-makers, processes and outputs. These elements interact through the 5 stage Health Policy Development Cycle (HPDC) (Figure 1).

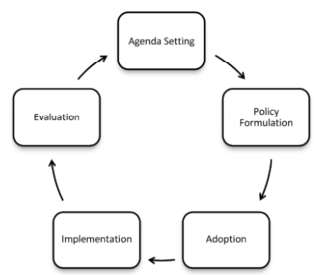


Figure 1 – Key Stages in the Health Policy Development Cycle [8]

Stage 1 – Agenda Setting: Identifies and defines the current policy issues and sets an agenda as to what problems need to be addressed [7, 8, 11].

Stage 2 – Policy Formulation: Sets policy boundaries between State and Federal health authorities and creates and/or changes actual policy content [8].

Stage 3 – Adoption: Sees the policy amendments brought to attention of policy approval bodies and put into force by State or Federal legislative measures [8, 12].

Stage 4 – Implementation: Action is taken to officially put approved policy changes into practice – more particularly formalising these actions by writing and documenting them into legislation.

Stage 5 – Evaluation: Involves the process of monitoring, analysing, critiquing and assessing existing or proposed policies. It is designed to help governments implement these policies in an effective and efficient manner through heavy examination of their policy content and their effects [11, 12].

Tools

Policy developers use a range of tools during the various stages of the HPDC. Stakeholder Analysis is used during Stages 1 and 2 to analyse the behaviour, responsibilities, interests, agendas, and common resources or interactions between different stakeholders and how these impact the policy process [11, 12].

Stage 3 may include the use of policy evaluation software, such as Policy Maker. These types of software focus on analysing the political dimensions of adopting public policy and decision making around such changes. They can also assist stakeholders in determining the feasibility of the policy to be implemented. Outputs are in the form of graphs, statistical charts and opportunity/obstacle narratives [11].

Tools used in Stage 5 (Evaluation) include document analysis and stakeholder mapping. Document analysis involves a systematic review of existing policy documentation and determining the clarity and comprehensiveness of the internal statements in alignment to the reflection of intended outcomes [9]. Stakeholder analysis has also been applied to the latter stages of the policy process. By deconstructing political dimensions, linkages can be made to stakeholders and objectives particularly between the distinct micro, meso and macro policy levels [14]. Further analysis on the level of control each stakeholder has in relation to proposing new agenda items can also trigger recommencement of the policy development cycle.

Issues

The current tools and approaches to policy analysis present several limitations. In document analysis, there a disregard to other contributing factors which are generally involved with the policy process such as stakeholders, resources and monitoring processes. Application of stakeholder analysis/mapping within a healthcare context has also proven difficult in drawing proper conclusions in relation to policy assessment. With the continuous changes occurring within health environments, it is difficult to easily gather information which is "cross-sectional" or "static" to a particular period of time [13, 14]. The timeframe for which health policy data is collected and analysed can also affect the relevance of analysis that informs decision-making. When using computer-aided policy tools, the quality of the assumptions and judgments made by the analyst can affect the validity of the results produced by the application [11, 15, 16].

This suggests there is a major deficit between the formulation stage and the implementation stage [9] as those that formulate the policy are usually not involved in the implementation of resulting policy changes at the operational level [9, 17]. In addition, current methods and tools used for evaluating and analysing policy within particular stages of the policy cycle are isolated to a singular perspective potentially resulting in an incomplete understanding of workflow, impact and outcomes. This deficit can be overcome by introducing visual patient journey modelling to the early stages of the HPDC to provide evidence to policy developers and help them to understand the potential impact of policy changes on patient flow and service delivery at an operational level.

Research Setting & Design

This research study utilises a multi-method approach for visualising the impact of policy changes particularly in a community care setting. Participatory Action Research is used to gather healthcare stakeholder input. The aim of the research is to better understand service design in a community care setting, whilst also determining the impact of State-level policy interventions through the application of an emerging patient journey modelling technique known as Essomenic. Document analysis and healthcare provider focus groups also contributed to the development of the resulting models.

Research Setting

The study examines the process flow and service design of a six-week short-term post-discharge care program known as ComPacks (Community Packages). ComPacks is a non-clinical care program initiated by NSW Department of Health that promotes a safe transition of patients from hospital to home. Services include: meals on wheels, mobility and transport, assistive care and home nursing. The main focus of ComPacks is to reduce the probability of hospital readmissions caused by post-hospital complications, as well as provide an interconnecting pathway for on-going services for longer-term care and management. There are currently several community care sites across New South Wales that deliver ComPacks Services to eligible patients in their local communities. One such site, Liverpool Catholic Care is part of a major non-profit organisation, which is responsible for the delivery of 120 programs in areas such as ageing, disability care, youth and family, education, training and support services.

Participatory Action Research

Participatory Action Research (PAR) involves all relevant parties actively examining together, current action (which they experience as problematic) in order to change and improve it. Research is designed to address specific issues identified by local stakeholders, and the results are directly applied to the problems at hand [25]. Staff from the local ComPacks office were included in a number of group workshop sessions. They described the flow according to the old policy (pre) and as per the latest policy changes (post).

Patient Journey Modelling

Patient journey modeling (PJM) visually describes the journey of patients as they transition through the system of care [18], [19]. It is an improvement tool used to represent the healthcare system as a whole and identify the areas in which service delivery can be improved [18–24]. This study focused on the application of an emerging healthcare redesign technique known as Essomenic. Essomenic is a patient journey modelling technique that has had successful applications within the healthcare domain, in areas such as oncology, ambulatory care, indigenous maternity care and mental health [18–24]. Essomenic patient journey modelling, uses the principles of patient centered care and drives the redesign activity from a patient viewpoint identifying items of value to the patient and placing them at the forefront of the modeling process [18]. It incorporates additional factors that contribute to the overall quality and delivery of care including patient needs, policies and guidelines, staff roles, information needs and provides considerable measurement capabilities [18]. Modelling syntax puts patients, shown in red, at the top of the model with each type of staff role involved in the patient flow allocated a unique colour.

Other key syntax includes:

- Blue oblongs = processes
- Green documents = paper-based information
- White system icons = electronic information supported by ICT
- Pink documents = clinical guidelines/policies
- White metrics boxes = measurement criteria

By including care providers in the model development process and producing visual outputs, stakeholders exhibit higher degrees of understanding and engagement [25]. The models also provide evidence on the impact of policy change pre- and post-implementation of the amendments.

Facilitated Patient Journey Modelling Sessions

Throughout the duration of the study, several facilitated PJM sessions or workshops were conducted with the purpose of gathering first-hand evidence of perceived impacts made to ComPacks workflow and service design. These modelling sessions/workshops serve as an informal collaborative environment aimed to educate participants of the methods used (PJM), whilst also encouraging and empowering them to be part of the problem solving process. As was previously addressed, the workshops emphasise visually representing the personal experiences of ComPacks staff members during the delivery of ComPacks. The ComPacks staff members are also included in the validation process of the models produced. This is to ensure that models produced accurately depict workflow and service design of ComPacks services.

ComPacks Patient Journey Models

As compliance to policy is an important part of the ComPacks funding model, document analysis and modelling were undertaken to examine the impact of policy changes implemented in 2012. The ComPacks policy was analysed to produce two viewpoints.

1. Patient flow prior to policy changes made in late 2011. This model visually represents the normal course of events as per the previous (2005-2006) ComPacks policy.
2. Patient flow post-policy changes.

   This model visually represents the current patient flow as per the latest policy implemented in 2012. See Figure 2.

Production of models pre- and post-policy changes provides evidence as to the impact of policy changes on patient flow into, through and out of the ComPacks program, and highlights if service delivery processes have been improved as a result of the decreed State-level policy changes. The scope of each model begins with an in-hospital eligibility assessment until their exit from the ComPacks program. Figure 2 shows the first page of the post-policy change Essomenic patient journey model.

Comparative Analysis Between Pre and Post Policy Change Models

Comparative analysis of both models was conducted to determine how ComPacks was impacted by the dictated policy changes. The Essomenic modelling technique allows for the easy identification of factors that may potentially affect the overall service design and workflow of ComPacks. These factors include:

- Change in total number of processes
- Repetition of tasks or duplicated processes
- Introduction of new staff roles

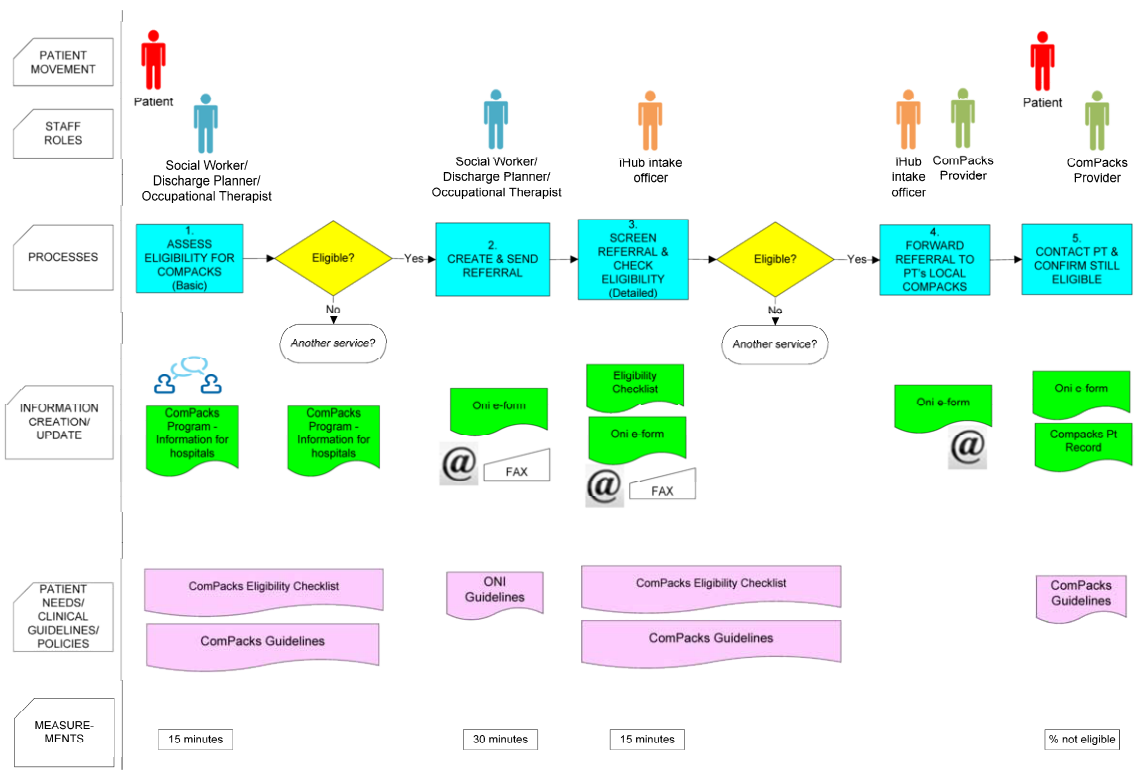


Figure 2 – Page 1 of post-policy change (Essomenic patient journey model)

Results and Discussion

Due to space restrictions the **pre-change policy model** has not been included in this paper however the description of an excerpt of the patient flow, **post-policy change** (see Figure 2) is as follows:

1. The in-hospital team discusses patient assistance requirements and performs a basic ComPacks eligibility check;
2. If eligible, the in-hospital team creates a referral and transmits this to Triple I Hub by email or fax (Triple I Hub is the centralised intake and intervention centre for the Local Health District);
3. The referral is screened for validity and a detailed eligibility assessment is performed by the Triple I Hub intake officer;
4. If eligible, the referral is forwarded to the patient's local ComPacks provider;
5. The ComPacks provider contacts the patient and confirms the patient is still eligible for an assistance package.

The journey continues on from this point and each individual service provider also checks the patients' eligibility for their particular service prior to commencing service delivery.

From the preliminary analysis conducted, two key policy impacts were identified between the pre and post change patient flow: the number of processing steps in the new policy has increased by over 44% and the number of service eligibility assessments have increased by 100% (see Table 1).

The increase in the number of processes is mainly due to the introduction of a centralised referral centre known as Triple I Hub. This centre receives all referrals from hospitals in the Local Health District, screens them for validity and performs a detailed eligibility check.

Table 1: Key preliminary findings of impact Pre- and Post-policy changes

Item	Pre-policy change	Post-policy change	% change
# of processes	18	26	44% increase
# of eligibility assessments	2	4	100% increase

The impact of the introduction of the Triple I Hub is a delay in the time taken to commence service delivery to the patient post-discharge and a number of processing duplications around eligibility assessment.

Eligibility is in fact checked four times in the new patient flow as shown in Figure 3.

Figure 3: Eligibility checkpoints as per latest policy changes

The visual representation of the patient flow readily identifies the increase in the number of processing steps, as each process is numbered.

The duplication in eligibility checking is also easily identifiable due to the decision points which are shown as yellow diamonds (See decisions between processes 1 and 2 and then processes 3 and 4 on Figure 2).

The Essomenic method of visual patient journey modelling provided ComPacks management with clear evidence on how they have been negatively impacted by State-level decreed policy changes at a operational patient flow level.

As a result of this study, it has been demonstrated that patient journey modelling can be used to visualise the impact of policy changes in community care services. There is high value in utilising Essomenic type models to illustrate suggested policy amendments prior to and after the implementation of policy changes.

Conclusion

Health policy plays an important role in the delivery of healthcare services. Policy guides providers during care processes whilst ensuring consistent service delivery and compliance to service criteria needed to secure government funding. Policies for care services are often amended by government bodies with the aim of enhancing and optimising services provided. However, it is perceived that changes made to policy can negatively impact workers, resources and workflow as well as the patient outcomes during service delivery. Therefore, it is imperative that all involved stakeholders understand and comprehend the impact of such changes and the ripple effects they may have on service delivery. This will not only allow for better decisions to be made on current and future policies but it will generate basic understanding for all those involved with the policy development process. In this study, focus was placed on utilising patient journey modelling, more specifically Essomenic, as a tool to visualise the impact of policy change on patient flow and service delivery in community care. There is value to be seen in using patient journey modelling as a tool to visualise impact of changes made to policy. It is a visual overview with representation of all elements that contribute to the policy process and provides solid evidence on the impact of changes on patient outcomes during service delivery.

Future work will investigate using Essomenic as a tool to evaluate and assess the impact of proposed policy changes to provide new value evidence to feed into the earlier stages of the health policy development cycle.

References

[1] M. D. Fine, "Coordinating health, extended care, and community support services: reforming aged care in Australia," *J. Aging Soc. Policy*, vol. 11, no. 1, pp. 67–90, 2000.

[2] L. A. Kemp, E. Harris, and E. J. Comino, "Changes in community nursing in Australia: 1995–2000," *J Adv Nurs*, vol. 49, no. 3, pp. 307–314, 2005.

[3] H. Keleher, "Community Care in Australia," *Home Health Care Manag. Pract.*, vol. 15, no. 5, pp. 367–374, 2003.

[4] A. M. Spehar, R. R. Campbell, C. Cherrie, P. Palacios, D. Scott, J. L. Baker, B. Bornstad, and J. Wolfson, "Seamless care: safe patient transitions from hospital to home," DTIC Document, 2005.

[5] S. Taylor, M. Foster, and J. Fleming, *Health care practice in Australia*. Oxford University Press, 2008.

[6] World Health Organisation, "Health Policy," 2014. [Online]. Available: http://www.who.int/topics/health_policy/en/.

[7] P. Dugdale, *Doing health policy in Australia*. Allen & Unwin, 2008.

[8] G. R. Palmer and S. D. Short, *Health care & public policy: an Australian analysis*. Macmillan Education AU, 2000.

[9] K. K. Cheung, M. Mirzaei, and S. Leeder, "Health policy analysis: a tool to evaluate in policy documents the alignment between policy statements and intended outcomes," *Aust. Heal. Rev.*, vol. 34, no. 4, pp. 405–413, 2010.

[10] M. Minkler and N. Wallerstein, *Community-based participatory research for health: From process to outcomes*. John Wiley & Sons, 2010.

[11]G. Walt, J. Shiffman, H. Schneider, S. F. Murray, R. Brugha, and L. Gilson, "'Doing'health policy analysis: methodological and conceptual reflections and challenges," *Health Policy Plan.*, vol. 23, no. 5, pp. 308–317, 2008.

[12] S. Barraclough and H. Gardner, "A problem-oriented approach to health policy analysis," *Anal. Heal. Policy A Probl. Approach*, p. 1, 2007.

[13] R. Brugha and Z. Varvasovszky, "Stakeholder analysis: a review," *Health Policy Plan.*, vol. 15, no. 3, pp. 239–246, 2000.

[14] M. H. R. Mehrizi, F. Ghasemzadeh, and J. Molas-Gallart, "Stakeholder mapping as an assessment framework for policy implementation," *Evaluation*, vol. 15, no. 4, pp. 427–444, 2009.

[15] M. Reich, "Policy Maker - Computer-Assisted Political Analysis," 2012. [Online]. Available: http://www.polimap.com/poliwhat.html.

[16] A. Glassman and K. Buse, "Politics, and Public Health Policy Reform," *International Encylopedia of Public Health*. pp. 163–170, 2008.

[17] M. Exworthy, "Policy to tackle the social determinants of health: using conceptual models to understand the policy process," *Health Policy Plan.*, vol. 23, no. 5, pp. 318–327, 2008.

[18] J. Curry, "A Meta-Methodology for the Design and Modelling of Patient Journeys. School of Computing and Mathematics. University of Western Sydney. Sydney, Doctor of Philosophy, 2007," *Comput. Sci.*, vol. 168, p. 369, 2008.

[19] D. I. Ben-Tovim, M. L. Dougherty, T. J. O'Connell, and K. M. McGrath, "Patient journeys: the process of clinical redesign," *Med. J. Aust.*, vol. 188, no. 6 Suppl, pp. S14–7, 2008.

[20] M. Crepaz and J. M. Curry, "Improving the cancer patient journey," in *Health Informatics: Digital Health Service Delivery-The Future is Now!: Selected Papers from the 21st Australian National Health Informatics Conference (HIC 2013)*, 2013, vol. 188, p. 14.

[21] J. M. Curry, J. A. Fitzgerald, and K. Eljiz, "Patient journey modelling: using students as a vehicle for the diffusion of innovation," *Stud. Health Technol. Inform.*, vol. 168, p. 35, 2011.

[22] V. K. C. Manchaiah, D. Stephens, and R. Meredith, "The patient journey of adults with hearing impairment: The patients' views," *Clin. Otolaryngol.*, vol. 36, no. 3, pp. 227–234, 2011.

[23] A. Wales, "Managing knowledge to support the patient journey in NHS Scotland: strategic vision and practical reality," *Heal. Inf. Libr. J.*, vol. 22, no. 2, pp. 83–95, 2005.

[24] J. Curry, A. Fitzgerald, A. Prodan, A. Dadich, and T. Sloan, "Combining patient journey modelling and visual multi-agent computer simulation: A framework to improving knowledge translation in a healthcare environment," *Stud. Health Technol. Inform.*, vol. 204, pp. 25–31, 2014.

[25] J. Curry and A. Prodan, "Using visual tools to improve clinical engagement and group understanding of complex IT concepts.," *Stud. Health Technol. Inform.*, vol. 192, p. 1028, 2013.

Address for correspondence

Rosemary Sayvong
University of Western Sydney
School of Computing, Engineering and Mathematics,
Telehealth Research and Innovation Laboratory (THRIL)
16779554@student.uws.edu.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-434

Challenges and Hurdles of eHealth Implementation in Developing Countries

Mandirola Brieux HF[a], Bhuiyan Masud JH[b], Kumar Meher S[c], Kumar V[c], Portilla F[d], Indarte S[e], Luna D[a], Otero C[a], Otero P[a], González Bernaldo de Quirós F[a]

[a] *Hospital Italiano, Buenos Aires, Argentina*
[b] *UNICEF, Bangladesh*
[c] *All India Institute of Medical Sciences, Delhi, India*
[d] *HL7 Colombia, Bogota, Colombia*
[e] *SUEIIDISS, Montevideo, Uruguay*

Abstract

Health informatics has the potential to improve the security and quality of patient care, but its impact has varied between developed and developing countries. Related to this, the objective of this study is to identify the challenges and hurdles to improve eHealth in developing countries. We surveyed experts to discover their opinions about five general questions: economic support by Government for eHealth, Government education or training projects in the field, issues related to cultural or educational problems for the implementation of eHealth, policies in terminology or messaging standards and eHealth status policies for long periods. The respondents answered affirmatively in these proportions: 1. Economic support policies 58%, 2. Training policies 25%, 3. Cultural and educational problems 95%, 4. Standards policies, 38%, 5. Policies for long periods, 50% Conclusion: Our survey has shown that the important problems that need to be addressed in order to implement e-Health in developing countries are firstly, cultural and educational, secondly, economic resources and thirdly policies for long periods.

Keywords:

Computer system; eHealth; Medical Informatics; Cultural; Educational.

Introduction

eHealth is the transfer of health resources and health care by electronic means. It involves three main parts: the delivery of health information, for health professionals and health consumers, through the Internet or telecommunications. It is using the power of Information Technology (IT) and e-commerce to improve public health services, e.g. through the education and training of health workers [1].

eHealth has a great importance in the management of health care services. There is no doubt about the advantages of information technology applied to health [2], but in most developing countries there are serious barriers to its effective implementation [3]. Information technology may allow significant improvements across various aspects and has the potential to benefit both developed and developing countries. The World Health Organization (WHO) identified the use of eHealth as a priority skill in the development of human resources in health (human resources in eHealth requires people with knowledge in medical informatic and standard terminology (e.g., ICD-10, SNOMED and HL7 messaging standards.)) Furthermore, it is increasingly recognized as a crucial piece to improve health systems to achieve the WHO Millennium Development Goals [4].

Poor strategic planning and a lack of international standards consume government budgets without reaching good results [5]. Beginning to develop systems without having a defined framework means implementations might fall into serious common mistakes. Without first identifying standards policies, network connectivity and Internet access, master files and unique identifiers it could lead us to waste time and resources. Even though experience and resources for early implementation projects in developed countries are available, in many developing countries there are still barriers and difficulties to access these resources.

eHealth care is a challenge that all countries face today, irrespective of their development status [4]. Some aspects that threaten system implementation in the health sector involve economic resources [6], income disparities, exorbitant costs of usage fees, excessive costs for even rudimentary health information systems [7], lack of human trained resources [8], lack of governmental policies that address a well-defined health system that incorporates eHealth [9], cultural aspects [10] and some resistance to the use of computers for health care processes. Also, standards policies have the potential to play a major role in system interoperability. Ensuring standards, generating guidelines, and introducing essential policies based on effective and efficient evidence could be necessary [11].

The aim of this work is to identify the main challenges and hurdles for the eHealth maturity of economically developing countries, located in Asia and South America.

Materials and Methods

In order to collect data about challenges and hurdles for eHealth implementations, we conducted a survey. A short, semi-structured, confidential interview was conducted by one physician trained in the field with special care to query respondents to answer about their spontaneous perception or personal knowledge.

Setting: Data collection was carried out among those attending INFOLAC 2014 the Latin American Conference on Medical Informatics organized by the Uruguayan Society of Health Informatics in Montevideo Uruguay from 16th to17th October 2014 [12] and APAMI 2014, the Asia Pacific Association for Medical Informatics Conference organized by the Indian Association for Medical Informatics (IAMI) in New Delhi, India from 30th Oct to 2nd November 2014 [13].

Design: The design of the study is a cross-sectional descriptive study.

Sampling strategies: From the meeting attendees, consecutive convenience sampling was performed on

attendees until three participants from each country participating in the events had responded to the survey.

The number of participants proposed for this study was 60 surveys in total.

Inclusion criteria: Individuals carrying out activities in the field of medical informatics in their country attending either INFOLAC 2014 or APAMI 2014.

Exclusion criteria: Those who could not answer more than two questions will be excluded.

Attendees will be able to add comments and suggestions that will be used in future studies.

To determine the obstacles to implementing eHealth in developing countries, we first identified the elements necessary to develop eHealth according to other studies, mainly in developing countries [14]. These variables are: 1) Economic support by the government for eHealth, 2) eHealth training, 3) Cultural or educational problems for implementation of eHealth, 4) Standards policies in terminology or messaging, 5) Polices in eHealth.

Questions for INFOLAC 2014 were made in Spanish and in English for APAMI 2014.

The questions for the anonymous survey were:

1. Is there any economic support by Government for eHealth?
2. Does Government organize eHealth courses or training?
3. Do you consider that there are cultural or educational problems for the implementation of eHealth?
4. Does your country have any standards policy in terminology or messaging in eHealth?
5. Does your country have eHealth status policies for long periods?

Results

All respondents agreed to participate in the study. Personal information was kept confidential. From the 60 proposed surveys, 27 were from South America and 33 from Asia. Four surveys were excluded due to not fitting the inclusion criteria.

The results of the survey showed the following:

1. "Is there any economic support by Government for eHealth?" Answer was 'Yes' in South America 33% of the time (Table 1 and Figure 1) and in Asia, 79% (Table 2 and Figure 2). The result across all countries was 58% (Table 3 and Figure 3).
2. "Does Government organize eHealth courses or training?" Answer was 'Yes' in South America 15% of the time (Table 1 and Figure 1) and in Asia 33% (Table 2, Figure 2). The result across all countries was 25% (Table 3 and Figure 3).
3. "Do you consider that there are cultural or educational problems for the implementation of eHealth?" In South America, respondents answered 'Yes' 100% of the time (Table 1 and Figure 1), and 91% in Asia (Table 2 and Figure 2). Considering all countries, 'Yes' was selected 95% of the time (Table 3 and Figure 3).
4. "Does your country have any standards policy in terminology or messaging in eHealth?" The answer was 'Yes' in South America 33% of the time (Table 1 and Figure 1), and in Asia 42% (Table 2 and Figure 2). There was a positive answer across all countries 38% of the time (Table 3, Figure 3).
5. "Does your country have eHealth status policies for long periods?" Answer was 'Yes' for this question in South America for 41% (Table 1, Figure 1) of respondents, and 58% in Asia (Table 2 and Figure 2). The results for all countries was 50% (Table 3, Figure 3).

Table 1 - South American respondents

			%
1	Economic support	9	33
2	Training	4	15
3	Cultural or educational problems	27	100
4	Standards policy	9	33
5	Policies for long periods	11	41
	Total	27	

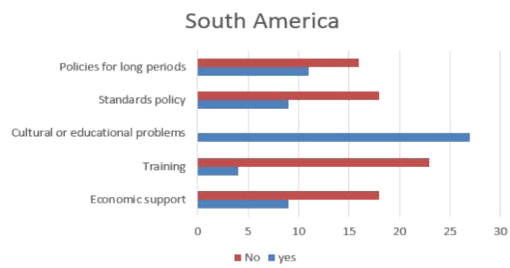


Figure 1 - South America respondents

Table 2 - Asia respondents

			%
1	Economic support	26	79
2	Training	11	33
3	Cultural or educational problems	30	91
4	Standards policy	14	42
5	Policies for long periods	19	58
	Total	33	

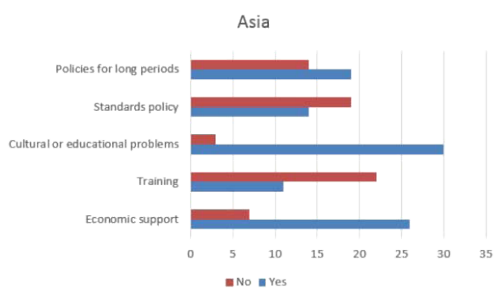


Figure 2 - Asia respondents

Table 3 - All respondents

			%
1	Economic support	35	58
2	Training	15	25
3	Cultural or educational problems	57	95
4	Standards policy	23	38
5	Policies for long periods	30	50
	Total	60	

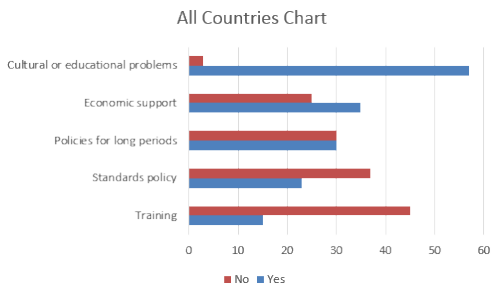


Figure 3 - All respondents

Table 4 - All respondents, comparison by region in %

	Asia	S.America
Economic support	79	33
Training	33	15
Cultural or educational problems	91	100
Standards policy	42	33
Policies for long periods	58	41

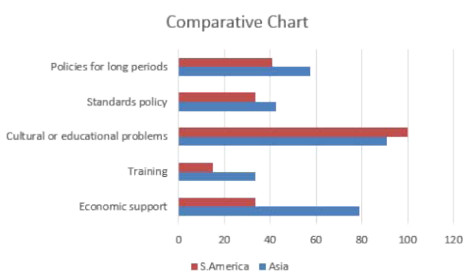


Figure 4 - All respondents, comparison by region

Results shown in both regions were very similar about the role of the states in relation to national policies for long projects, standards policies, cultural issues and education for eHealth. The role of the state in relation to training and economic support in eHealth was perceived differently in the analyzed regions, being more intense in Asia than in South America. (Table 4 and Figure 4).

Discussion

This was a preliminary survey to get an overview of the situation in our region and pave the way for more complex studies.

Training of human resources in eHealth is critical, because we think it is one of the limiting steps of high impact, not only from a technical standpoint, but as a change management tool.

It is interesting to note that participants agree that eHealth development will require more universal eHealth interoperability standards and strategies to overcome technical infrastructure barriers and address privacy, security, and other legal requirements [15]. We should take advantage of the lessons learned in developed countries in order to optimize the strategies to achieve this goal.

Concerning regulatory legal and policy framework, it is difficult, in most of our countries to find clear policies and coordination between state or governmental agencies and eHealth initiatives. This is a huge obstacle to implementing eHealth in developing countries.

Instability in politicals issues make it really difficult to find policy governance for long-term projects. Governments want centralized systems instead of dictating long-term policies that enable the local developers to interoperate with other regional developments through clear policies standards. In many of our countries there are no state policies, and if there are, they often change when the government and projects fall or are changed by others. Developing policies for eHealth requires long-term political times exceeding consensus and projects [16].

One of the limitations of this study is the sample. We collected the data at only two events in South America, during INFOLAC 2014 conference and during APAMI 2014. It could be necessary to extend the coverage of the sample for our next studies. Even so, our results are consistent with the work of Lacroix and colleagues that report that one of the main problems has to do with the educational training [17].

Conclusion

eHealth is a promising concept to achieve better health for all. The present study suggests there is a huge gap in cultural and educational issues regarding eHealth. Also, according to our results, we must take into consideration long-term eHealth projects. It is very important to consider interoperability and standards in eHealth as well. We have to work hard to solve the obstacles presented to achieve health for all and not settle for anything less.

Acknowledgments

The authors acknowledge Rachel Leidi who helped in the translation and correction of this work.

References

[1] WHO | E-Health [Internet]. WHO. [cited 2015 Mar 27]. Available from: http://www.who.int/trade/glossary/story021/en/

[2] Lasker RD, Humphreys BL, Braithwaite WR, Committee US--PHDC. Making a Powerful Connection: The Health of the Public and the National Information Infrastructure [Internet]. [cited 2014 Nov 27]. Available from: http://www.nlm.nih.gov/pubs/staffpubs/lo/makingpd.html

[3] Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a

focus on meaningful use. Ann Intern Med. 2014 Jan 7;160(1):48–54.

[4] Merrell RC. Review of National e-Health Strategy Toolkit. Telemed E-Health. 2013 Oct 19;19(12):994–994.

[5] Sluijs MB, Veeken H, Overbeke AJPM. [Deficient information in developing countries: Internet alone is no solution]. Ned Tijdschr Geneeskd. 2006 Jun 17;150(24):1351–4.

[6] Chinnock P, Siegfried N, Clarke M. Is Evidence-Based Medicine Relevant to the Developing World? Evid Based Complement Alternat Med. 2005 Sep;2(3):321–4.

[7] Ashraf H. Countries need better information to receive development aid. Bull World Health Organ. 2005 Aug;83(8):565–6.

[8] Oak M. A review on barriers to implementing health informatics in developing countries. J Health Inform Dev Ctries [Internet]. 2007 Dec 18 [cited 2014 Nov 28];1(1). Available from: http://jhidc.org/index.php/jhidc/article/view/4

[9] Ahern DK, Kreslake JM, Phalen JM, Bock B. What Is eHealth (6): Perspectives on the Evolution of eHealth Research. J Med Internet Res [Internet]. 2006 Mar 31 [cited 2014 Nov 29];8(1). Available from: http://www.jmir.org/2006/1/e4/

[10] Lee MEPH, MD. The Strategy That Will Fix Health Care [Internet]. Harvard Business Review. [cited 2014 Nov 29]. Available from: https://hbr.org/2013/10/the-strategy-that-will-fix-health-care

[11] Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med. 2008 Apr 17;358(16):1732–7.

[12] VI Congreso Iberoamericano de Informática Médica, INFOLAC 2014 | Salud-e [Internet]. [cited 2014 Dec 19]. Available from: http://www.salud-e.cl/prensa/vi-congreso-iberoamericano-de-informatica-medica-infolac-2014/

[13] APAMI: Asia Pacific Association for Medical Informatics [Internet]. [cited 2014 Dec 19]. Available from: http://www.imia-medinfo.org/new2/node/157

[14] Wood A. Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom. World Bank Econ Rev. 1997 Jan 1;11(1):33–57.

[15] Ferrão LJ, Fernandes TH. Community oriented interprofessional health education in Mozambique: one student/one family program. Educ Health Abingdon Engl. 2014 Apr;27(1):103–5.

[16] Luna D, Almerares A, Mayan JC, González Bernaldo de Quirós F, Otero C. Health Informatics in Developing Countries: Going beyond Pilot Practices to Sustainable Implementations: A Review of the Current Challenges. Healthc Inform Res. 2014 Jan;20(1):3–10.

[17] Lacroix A, Lareng L, Padeken D, Nerlich M, Bracale M, Ogushi Y, et al. International Concerted Action on Collaboration in Telemedicine: Recommendations of the G-8 Global Healthcare Applications Subproject-4. Telemed J E Health. 2002 Jun 1;8(2):149–57.

Address for correspondence

humberto.mandirola@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-438

On the Development of a Hospital-Patient Web-Based Communication Tool: A Case Study From Norway

Conceição Granja[a,b], Kari Dyb[a], Stein Roald Bolle[a], Gunnar Hartvigsen[a,b]

[a] *Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[b] *Department of Computer Science, UIT – The Arctic University of Norway, Tromsø, Norway*

Abstract

Surgery cancellations are undesirable in hospital settings as they increase costs, reduce productivity and efficiency, and directly affect the patient. The problem of elective surgery cancellations in a North Norwegian University Hospital is addressed. Based on a three-step methodology conducted at the hospital, the preoperative planning process was modeled taking into consideration the narratives from different health professions. From the analysis of the generated process models, it is concluded that in order to develop a useful patient centered web-based communication tool, it is necessary to fully understand how hospitals plan and organize surgeries today. Moreover, process reengineering is required to generate a standard process that can serve as a tool for health ICT designers to define the requirements for a robust and useful system.

Keywords:

Elective surgery cancellations, pre-operative planning, web-based communication, process optimization, Norway.

Introduction

Considering the trends of healthcare, hospitals today have an extended focus on reducing costs [1]. Surgery is one of the most costly services provided by hospitals [2, 3]. However, elective surgeries are still regularly cancelled; in western countries, cancellation rates between 10 and 40 % have been reported [3-5]. Furthermore, up to 20 % of elective surgeries are cancelled on the day of surgery [6-8] and it is also identified that 50 % of these cancellations might be avoided [3, 9, 10]. Evidence [7, 11] points to lack of information as being the main cause for elective surgery cancellations, referring to information that existed but was not available when required for surgery planning. In line with what is reported in the literature, our site of research, the University Hospital of North Norway (UNN), has identified that 50 % of all cancellations are related to inadequate planning due to lack of information [12]. In the described health scenario, health ICT emerges as part of a discourse that emphasizes cost-effectiveness and active involvement of patients as a policy to meet the growing demand of care and the expected shortage of health care professionals [1].

The aim of the research project "eTeam-Surgery" is to reduce the number of elective surgery cancellations at UNN. In today's surgical process, the information required for anesthetic evaluation is gathered after the patient admission, based on a health assessment questionnaire. eTeam-Surgery will provide a digitalized tool for two way communication between the hospital and the patient prior to hospital admission. Such a tool will enable the hospital to collect the missing information from the patient at home and at an earlier stage in the pre-operative planning process.

The development of an efficient and functional web-based tool for hospital-patient collaboration is not an easy task, and it has not always been successful. As the development of health ICT grows, there are also an increasing number of reports on unsuccessful implementation projects, challenges and unforeseen consequences of ICT in health care, particularly in hospitals [13-24]. A contributing factor to such results may be found on the focus of health ICT on improving individual tasks rather than supporting value added care processes. By supporting individual tasks, ICT is focusing on the provider. This is a significant contribution to a lower quality and high cost health care. On the other hand, process focused care is centered on the patient. It integrates the teamwork (e.g. patients, physicians, nurses, caregivers, managers, and administrative personnel) to provide high quality and efficient care throughout the process. Value added care processes are the goal of the patient centered health care. However, few health care processes have been modeled comprehensively enough to provide a basis for specifying software requirements to health ICT designers. Thus, health ICT designers have focused on supporting the work of individual care team members by taking existing paper-based tools, as their models. The result is that most health ICT systems do little to support care teams. Hence, prior to development, eTeam-Surgery carried out an in-depth study of the pre-operative planning at UNN.

In this paper, is described an in-depth study of the pre-operative planning at UNN. The aim of this study was to describe and understand the process to define the requirements for a hospital-patient web-based communication tool. The paper is divided in four sections. In the first section, the health ICT limitations are introduced, and the aim of the study is described. In the second section, a brief introduction to the existing process modeling tools is presented. The methodology to model health care processes is described and explained in the third section. In the last section, discussion and conclusions, the authors elaborate on the need to understand care processes in order to define the requirements for a hospital-patient web-based communication tool.

Background

Workflow modeling is the basis for process optimization, and exists in a wide range of modeling languages. Graphical modeling languages can be divided in languages oriented to: (a) information flow, (b) control flow, or (c) objects [25]. In this work the modeling languages applied to the control flow are considered for a better accommodation of health care specific workflow.

Petri nets are a formalism of graphical modeling, with application in systems description and in the study of information processing, characterized by being asynchronous, parallel, non-deterministic and/or stochastic [26]. This formalism was applied to several problems in the healthcare field, such as the evaluation of patient monitoring systems [27, 28], the modeling of patient flows in progressive systems [29-31] or the modeling of logistic processes in hospitals [32].

Apart from the Petri nets, process chains oriented to operations, denoted Event-driven Process Chains (EPC) is another modeling formalism widely used in the healthcare field. EPC is presented as a modeling concept of logical and time dependencies of processes, and was applied to the information flow management between the different systems [33].

Materials and Methods

The eTeam-Surgery project group has developed a three-step methodology to gather the knowledge required to model the pre-operative planning process. In order to keep this paper self-contained, the empirical methodology will be described briefly hereafter, for further information please refer to [34].

1. Gather data on the hospital's representation of the elective surgery cancellation problem;
2. Observations and interviews at the hospital, related to the pre-operative planning processes at the department level;
3. Individual, in-depth interviews with all professional groups involved in pre-operative planning at a specific hospital department.

In Stage 1, the aim was to gather knowledge on UNN's understanding of the elective surgical cancellation problem, and the hospital representation of the pre-operative planning process. One document, containing information on the use of resources involved in surgery at the hospital was identified and studied [12]. In 2012, UNN initiated a Lean project in order to optimize the elective surgical process. Researchers from the eTeam-Surgery group followed this project.

In Stage 2, the pre-operative planning process at different departments at UNN was investigated. This comprised three weeks of fieldwork at the Surgery and Intensive care clinic, doing interviews while following an anesthesiologist and an anesthetic nurse. In addition, thirteen interviews with physicians, nurses and administrative personnel were conducted, at six different departments. The interviews were semi-structured, done at the workplace, and lasted between thirty minutes to two hours.

During the first two stages, two departments were described to be more efficient. However, these departments still evidenced a representative number of cancellations. One of the departments was chosen to proceed with an in-depth study in Stage 3. The chosen department is not revealed due to ethical reasons. In Stage 3, representatives from all the professional groups involved in the pre-operative planning process at UNN were addressed. At this specific department, extensive knowledge on the pre-operative planning process was collected. The department-specific interviews were semi-structured, conducted at the workplace, and lasted between one to two hours.

Results

The observations of, and interviews with, the health professionals at the chosen department at UNN, described in the previous section, allowed the definition and mapping of the generic pre-operative process model shown in Figure 1 and Figure 2. (For further information on the interviews, refer to [35].) A process model facilitates a systematic description of the events permitting the identification of each decision activity and the health worker responsible for that activity. In addition, it allows us to learn about the information flow, and allows the opportunity to identify the underlying processing issues that may be compromising the availability of vital patient assessment information.

At UNN, as seen in Figure 1 and Figure 2, the final pre-operative planning is often done after the patient has arrived for the scheduled surgery, which means, the final pre-operative planning might be completed the day prior to, or even on the day of the surgery. During this final planning process, new information, which may lead to cancellations, is gathered from the patients. Figure 1 and Figure 2 show additional steps when supplementary / new information might be necessary. The first step is in the beginning of the pre-operative planning process during the referral evaluation. A descriptive comparison of the referral evaluation process in both figures also evidences the ambiguity of the resources' role on the process. For example, as depicted in Figure 1, the task of quality assessment and subsequent referral, or request for additional information, is performed by the secretary. Figure 1 also portrays identified telephone calls between the surgeon and the patient as an additional alternative to collect lacking information. None of these tasks are described in Figure 2.

Discussion and Conclusions

As explained in the Introduction section, Health ICT orientation to individual tasks reflect the focus of health care itself: The majority of clinical departments behave as discrete and independent sets of physicians, nurses, and other health personnel instead of a single team [36]. This is partly due to the autonomy of most clinical departments.

eTeam-Surgery started by mapping and evaluating the pre-operative process at UNN, and explored a system for gathering information from a patient on his/her condition through a personal health assessment questionnaire. This process facilitated the evaluation as to whether if, and how, part of the process could be migrated from the hospital to the patient at home through electronic collaboration.

The comparison of the pre-operative planning process models, shown in Figure 1 and Figure 2, evidences the heterogeneity of work patterns at the individual level. The narratives of the different health professions describe different patient pathways for the same process. This analysis of hospital workflow evidences the superfluous work required by hospital personnel to develop and document the information required for the patient to move forward in the EHR pathway. In such an environment, it is very difficult to develop a Health ICT system that both supports the full process and is useful for each member of the care team. Therefore, we argue that in order to develop a useful patient centered web-based communication tool, it is imperative that the processes by which hospitals plan and organize surgeries are fully understood, and that process reengineering be employed in the generation of standard processes of robust Health ICT system.

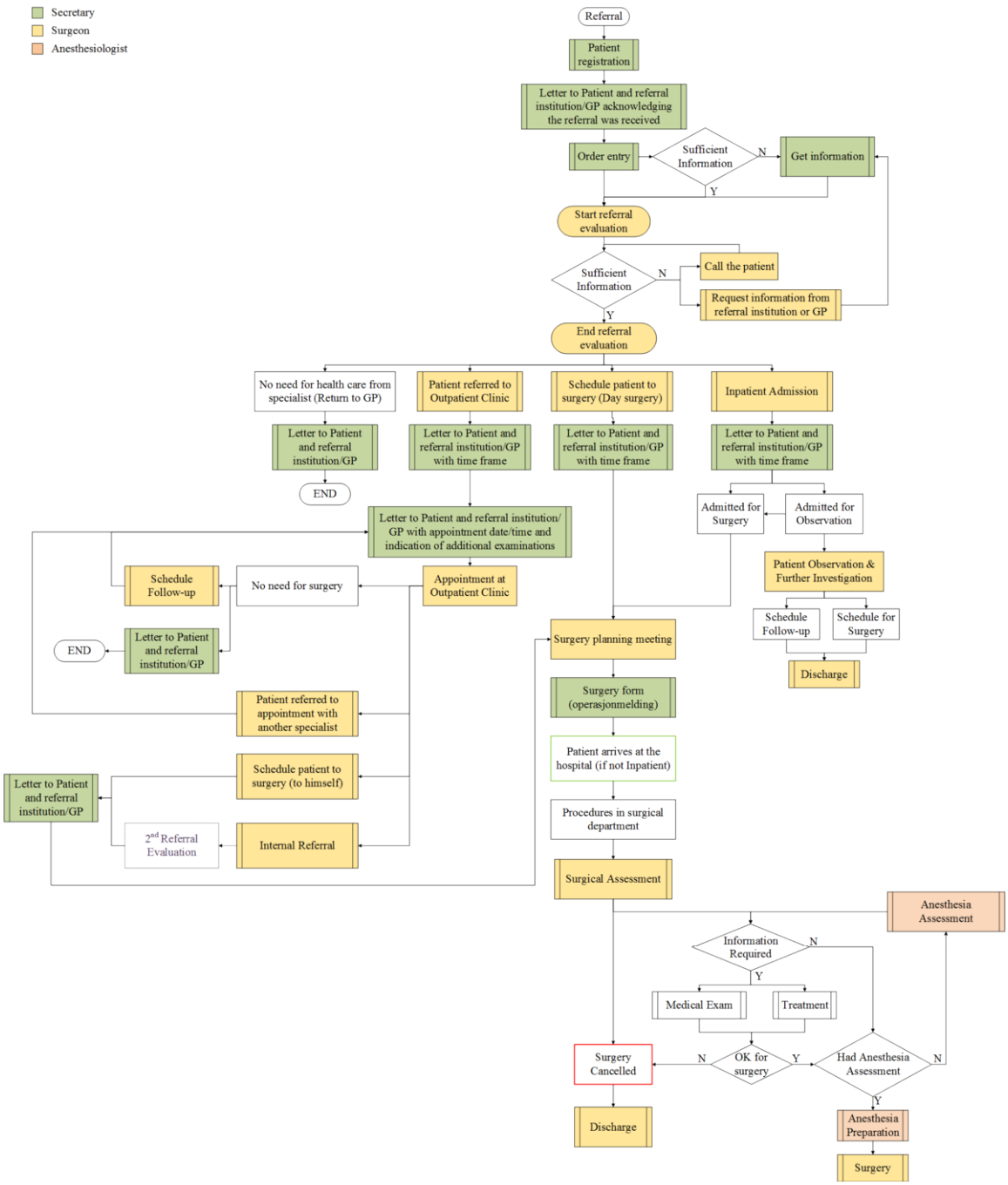


Figure 1 - Pre-operative planning process as experienced by surgeons and physicians.

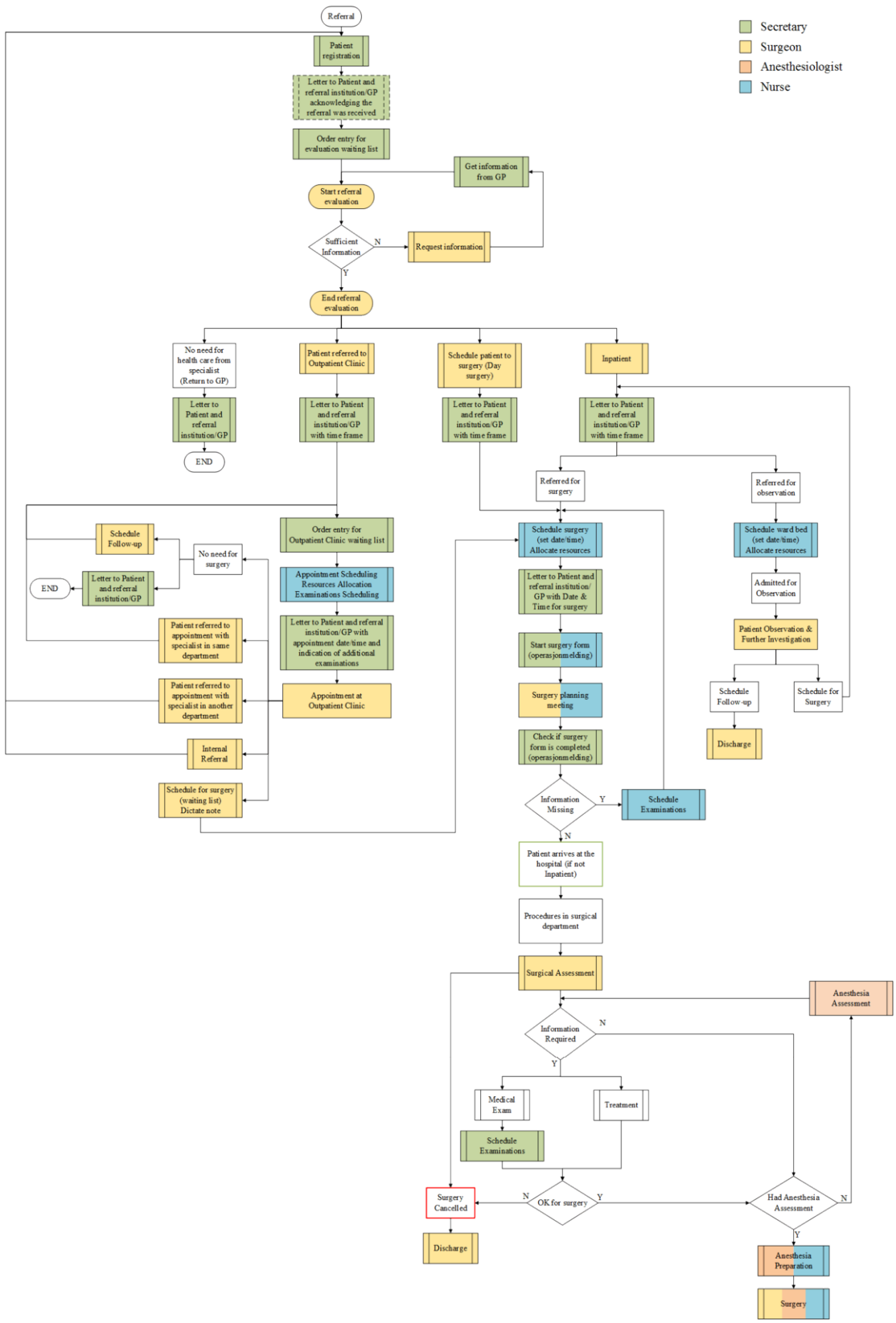


Figure 2 - Pre-operative planning process as experienced by coordinating nurse.

References

[1] Oudshoorn N. Physical and digital proximity: Emerging ways of health care in face-to-face and telemonitoring of heart-failure patients. Sociology of health & illness 2009: 31: 390-405.

[2] Denton B, Viapiano J, and Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 2007: 10: 13-24.

[3] Schofield WN, Rubin GL, Piza M, Lai YY, Sindhusake D, Fearnside MR*, et al.* Cancellation of operations on the day of intended surgery at a major australian referral hospital. Medical Journal of Australia 2005: 182: 612-615 (in English).

[4] Foster A. Operating theatres: Review of national findings. Audit Commission for Local Authorities and the National Health Service, London, 2003.

[5] Ivarsson B, Kimblad PO, Sjoberg T, and Larsson S. Patient reactions to cancelled or postponed heart operations. J Nurs Manag 2002: 10: 75-81.

[6] Aaserud M, MariTrommald, and Boynton J. Elektiv kirurgi - strykninger, skjerming og effektivitet. Tidsskr Nor Lægeforen 2001: 21: 2516-2519 (in Norwegian).

[7] González-Arévalo A, Gómez-Arnau JI, DelaCruz FJ, Marzal JM, Ramírez S, Corral EM*, et al.* Causes for cancellation of elective surgical procedures in a spanish general hospital. Anaesthesia 2009: 64: 487-493.

[8] Yoon SZ, Lees SI, Lee HW, Lim HJ, Yoon SM, and Chang SH. The effect of increasing operating room capacity on day-of-surgery cancellation, Edgecliff, AUSTRALIE: Anaesthesia Society of Anaesthetists, 2009.

[9] Sanjay P, Dodds A, Miller E, Arumugam PJ, and Woodward A. Cancelled elective operations: An observational study from a district general hospital. J Health Organ Manag 2007: 21: 54-58.

[10] Trentman TL, Mueller JT, Fassett SL, Dormer CL, and Weinmeister KP. Day of surgery cancellations in a tertiary care hospital: A one year review. Journal of Anesthesia & Clinical Research 2010: 1.

[11] Knox M, Myers E, Wilson I, and Hurley M. The impact of pre-operative assessment clinics on elective surgical case cancellations. Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland 2009: 7: 76-78.

[12] Busund R. Rapport fra prosjekt: Optimal ressursutnyttelse av opperasjonskapasiteten i unn. Universitetssykehuset Nord-Norge, Norway, 2008 (in Norwegian).

[13] Starling J and Foley S. From pilot to permanent service: Ten years of paediatric telepsychiatry. Journal of Telemedicine and Telecare 2006: 12: 80-82.

[14] Whitten P, Holtz B, and Nguyen L. Keys to a successful and sustainable telemedicine program. International journal of technology assessment in health care 2010: 26: 211-216.

[15] Zanaboni P and Wootton R. Adoption of telemedicine: From pilot stage to routine delivery. BMC medical informatics and decision making 2012: 12: 1.

[16] Berg M. Implementing information systems in health care organizations: Myths and challenges. International journal of medical informatics 2001: 64: 143-156.

[17] Heeks R. Health information systems: Failure, success and improvisation. International journal of medical informatics 2006: 75: 125-137.

[18] May C, Mort M, Mair FS, and Finch T. Telemedicine and the future patient: Risk, governance and innovation, Economic and Social Research Council, 2005.

[19] May C and Ellis NT. When protocols fail: Technical evaluation, biomedical knowledge, and the social production of 'facts' about a telemedicine clinic. Soc Sci Med 2001: 53: 989-1002 (in eng).

[20] Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, and Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. International Journal of Medical Informatics 2012: 81: 746-760.

[21] KS. Ikt i helse- og omsorg 2008-20012 - strategi- og handlingsplan. Oslo, 2008 (in Norwegian).

[22] Andreassen HK. What does an e-mail address add?-doing health and technology at home. Social Science & Medicine 2011: 72: 521-528.

[23] Schreurs N. Fiasko eller fremtid? Computerworld. 2012 (in Norwegian).

[24] Wyatt JC and Sullivan F. Ehealth and the future: Promise or peril? BMJ 2005: 331: 1391-1393.

[25] Gadatsch A. Grundkurs geschäftsprozess-management, Wiesbaden, Germany: Vieweg+Teubner, 2010.

[26] Bobbio A. System modelling with petri nets. In: Systems reliability assessment. Springer, 1990: 103-143.

[27] Nyman MA, "Patient flow: Reducing delay in healthcare delivery," in *Mayo Clinic Proceedings* vol. 82, ed: Elsevier, 2007, p. 388.

[28] Cayirli T and Veral E. Outpatient scheduling in health care: A review of literature. Production and Operations Management 2003: 12: 519-549.

[29] Bailey NT. A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. Journal of the Royal Statistical Society. Series B (Methodological) 1952: 185-199.

[30] Pham D-N and Klinkert A. Surgical case scheduling as a generalized job shop scheduling problem. European Journal of Operational Research 2008: 185: 1011-1025.

[31] Cardoen B, Demeulemeester E, and Beliën J. Optimizing a multiple objective surgical case sequencing problem. International Journal of Production Economics 2009: 119: 354-366.

[32] Hans EW, Van Houdenhoven M, and Hulshof PJ. A framework for healthcare planning and control. In: Handbook of healthcare system scheduling. Springer, 2012: 303-320.

[33] Lamiri M, Xie X, Dolgui A, and Grimaud F. A stochastic model for operating room planning with elective and emergency demand for surgery. European Journal of Operational Research 2008: 185: 1026-1037.

[34] Granja C, Dyb K, Larsen E, Bolle SR, and Hartvigsen G. Methodology for health care process modelling: Bringing the health care complexity into health it system development. In: Scandinavian Conference on Health Informatics. Grimstad, 2014, pp. 17-21.

[35] Dyb K, Granja C, Bolle SR, and Hartvigsen G. Online patients in an offline health care sector: Are hospitals ready for electronic communication with patients? In: The Seventh International Conference on eHealth, Telemedicine, and Social Medicine. Lisbon, 2015, pp. 26-30.

[36] Knox GE and Simpson KR. Teamwork: The fundamental building block of high-reliability organizations and patient safety. Patient safety handbook 2004: 379-414.

Address for correspondence

Conceição Granja

University Hospital of North Norway, Norwegian Centre for Integrated Care and Telemedicine Sykehusvn. 23, 9019 Tromsø, Norway

E-mail: conceicao.granja@telemed.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-443

Barriers and facilitators to the introduction of digital pathology for diagnostic work

Rebecca Randell[a], Roy A. Ruddle[b], Darren Treanor[c,d]

[a] *School of Healthcare, University of Leeds, Leeds, UK*
[b] *School of Computing, University of Leeds, Leeds, UK*
[c] *Department of Histopathology, Leeds Teaching Hospitals NHS Trust, Leeds, UK*
[d] *Leeds Institute of Cancer & Pathology, University of Leeds, Leeds, UK*

Abstract

Cellular pathologists are doctors who diagnose disease by using a microscope to examine glass slides containing thin sections of human tissue. These slides can be digitised and viewed on a computer, promising benefits in both efficiency and safety. Despite this, uptake of digital pathology for diagnostic work has been slow, with use largely restricted to second opinions, education, and external quality assessment schemes. To understand the barriers and facilitators to the introduction of digital pathology, we have undertaken an interview study with nine consultant pathologists. Interviewees were able to identify a range of potential benefits of digital pathology, with a particular emphasis on easier access to slides. Amongst the barriers to use, a key concern was lack of familiarity, not only in terms of becoming familiar with the technology but learning how to adjust their diagnostic skills to this new medium. The findings emphasise the need to ensure adequate training and support and the potential benefit of allowing parallel use of glass slides and digital while pathologists are on the learning curve.

Keywords:

Informatics; Pathology; Microscopy; Qualitative Research; Learning Curve.

Introduction

While traditionally health informatics focused on the consulting room and the use of electronic patient records (EPRs) and computerised decision support systems, recent years have shown an interest in broader range of settings. Not only is the spread of EPRs and mobile technologies into secondary care leading health informatics researchers to explore ward settings, but we are gradually learning more about those areas of medicine whose work patients may not see but which play a key role in diagnosis, determining treatment, and assessing response to treatment, such as the work of clinical pathology laboratories [1]. This has led to studies of diagnosis around imaging systems, such as use of Picture Archiving and Communication Systems (PACS) in radiology [2,3]. At the same time, we are seeing interest in a broader range of technologies for diagnostic imaging, including increasingly pervasive wireless and sensor-based technologies [4,5]. In this paper, we turn to the underexplored area of cellular pathology: the diagnosis of disease through microscopic examination of tissue.

Cellular pathologists diagnose cancer and other diseases using a microscope to examine glass slides containing thin sections of human tissue. The tissue will have been stained using a chemical or immunologically based method. Haematoxylin and Eosin (H&E) is the most common stain and it highlights the most significant tissue structures well, while special stains use histochemical reactions to identify specific tissue components or organisms, that are less easily identifiable with an H&E stain.

It is now possible to digitise glass slides so that they can be viewed on a computer; the field of digital pathology is concerned with the development and evaluation of technologies that support this. Digital pathology promises a number of benefits, both in terms of efficiency and safety. Advocates of digital pathology highlight the potential for improved workflow; a digital system would allow the pathologist to be alerted to when new cases are ready to be viewed, as well as allowing the pooling of cases, resulting in a revolution of the workflow similar to that seen in radiology [6]. There is also the ease of obtaining second opinions electronically from a national or international source rather than having to send the glass slides through the post with the delay and the risk that they will get broken or lost in transit [7]. Slides can be simultaneously sent to several people for second opinions, which is not possible with glass slides. There is the reduced risk of getting slides mixed up so that a patient receives the wrong diagnosis, something that happens rarely but can have devastating consequences when it does [8], and there is the option to integrate decision support technology [9, 10].

Despite this potential, uptake of digital pathology for diagnostic work has been slow. While there have been positive reports about the use of digital pathology within education and training [11,12] and for teleconsultation [10], in relation to routine diagnostic work, research suggests scepticism and uncertainty amongst pathologists [13,14]. In a recent survey conducted in the United States, while 59% of respondents agreed that the benefits of digital pathology outweigh concerns, 78% perceived digital slides as currently being too slow to view for use in routine diagnostic work [15]. This fits with the findings of experimental studies which have found that it takes significantly longer to make a diagnosis on a digital slide compared to a glass slide [16,17]. A study of barriers and facilitators to the use of digital images in clinical practice that involved interviews with two radiologists and three pathologists, found a key barrier in pathology to be the perception that diagnostic performance is inferior with digital slides [9].

As part of a larger study concerned with the design and evaluation of a digital microscope for use in routine diagnostic

work [18-21], we undertook an interview study to better understand the barriers and facilitators to the introduction of digital pathology. In this paper, we present the findings of that study and discuss the implications of these findings for the implementation of digital pathology.

Methods

Data collection

Semi-structured, face-to-face interviews were conducted with nine consultant pathologists within our institution, a large teaching hospital. The interviewees were from a range of subspecialties and had varying levels of experience of digital pathology: two had almost no experience of digital pathology, three had used it for External Quality Assessment (EQA) schemes, one had used it for obtaining a second opinion, and two had experience of using it for research and/or teaching.

Interview questions explored what participants saw as the benefits and limitations of the conventional microscope and current workflow, benefits, limitations, anticipated impacts of digital pathology, and willingness to move to digital pathology. All interviews were audio recorded and later transcribed.

Analysis

An iterative approach was taken to data collection and analysis, to allow the collection of further data on emerging themes in subsequent interviews. Anonymised transcripts were entered in NVivo, software for qualitative data analysis. Data was analysed using thematically [22], with codes developed inductively.

Results

Here we present the findings from the interviews organised according to key themes. While interviewees discussed the benefits of digital pathology for multidisciplinary team meetings and teaching, we focus on attitudes towards digital pathology in the context of diagnostic work.

Benefits of digital pathology

The most commonly mentioned advantage of digital pathology was the ease of sharing digital slides in order to get second opinions, mentioned by eight of the nine interviewees:

'I have sent slides to others for second opinion and it works brilliantly. First of all we don't have to pack slides. Recently I had to send a case to Boston [...] and I didn't have to pack any slides. The main problem was packing slides across to America, it's so difficult now. So I scanned the slides […] and sent it across and within half an hour I got the reply. So he emailed me saying yeah the report I think it's fine. And the thing is we don't choose any particular area as when you take images and send it to another person […] This is you're sending the entire slide as you would be seeing down the microscope. So that is a big advantage.' (HC6)

While most comments were about the ability to share slides with other sites, a couple of interviewees also mentioned benefits in terms of being able to look at slides with colleagues:

'When you're getting a second opinion, it's much much easier if you're all looking at a screen, you can all point at it and know you're all looking at the same thing. Um, and that you can all point at it whether you're steering or not, whereas on this [microscope] the only person who can point really very easily is the person steering it.' (HC5)

Interviewees were also very positive about the idea of having remote access to slides, both for on call work but also to allow more general flexibility:

'The scope for it is for the very specialised stuff, if you're looking to get cover for cases around the country, that sort of thing, that's the one diagnostic role I can see.' (HC3)

Another perceived benefit, mentioned by four of the interviewees, is the ease of access to slides, so that previous slides can easily be viewed:

'The nice thing about electronic images is that you can store them, you can access them rapidly, so if I needed to go back to look at something I don't have to go back to the file to get it out. In theory. If there's a secure server, I could go straight to it and find the one that I want with some rapidity. […] that would be grand. […] it would allow laboratories to store large amounts of glass off site. Rather than having to keep it all on the premises. So that people can review stuff very quickly, scanned images, and very rarely have to go back to the original bit of glass, which is quite labour intensive.' (HC4)

The interviewees mentioned the benefit of being able to have multiple people using the slide at the same time, mentioned by one interviewee when discussing EQAs but also by two interviewees in the context of diagnostic work:

'I suppose now I have to complete a case and then give it to a colleague because then as long as a colleague is working on it I can't look at it, so I suppose that would be an advantage, […] it's more easy to share cases.' (HC7)

Two interviewees mentioned the durability of digital slides, that they do not get broken or lost, while also mentioned by two interviewees was the removal of 'glass work':

'You don't have all the glass work, filing, you don't have to constantly file loads of glass work and pull it out and refile it and so on.' (HC2)

Two interviewees also mentioned that the ability to have an overview of the whole slide which is not possible with the microscope:

'I like the way it can be navigated where you can see in a wider view the entire biopsy or pieces of tissue that you see on the glass slide. That is definitely an advantage.' (HC6)

Limitations of digital pathology

The most commonly mentioned disadvantage, mentioned by three of the interviewees, was the lack of familiarity with digital slides:

'It just takes a bit of getting used to, open it up, zooming in and out and going round. And sitting at a computer screen as opposed to looking down a microscope.' (HC2)

One interviewee felt that it was not just a case of learning to use the technology but also adapting her diagnostic skills:

'I can't imagine someone who's so used to this as me would want to change for my routine work to a digital way of reporting. [...] Because, yeah, I'd be relearning a lot of diagnostic skills but wouldn't have the time to do it. I wouldn't see any need to do it.' (HC3)

While the ability to share slides with other sites was seen as a benefit, one pathologist was concerned that digital slides could result in less interaction amongst pathologists within the department.

An issue which came up in the first two interviews and which we explored further in subsequent interviews was the impact digital pathology would have on the sense of 'immersion' that some pathologists report experiencing when working at the microscope. One of the interviewees felt that a move to digital pathology would have a negative impact on the sense of immersion. However, two of the interviewees were unsure whether or not the loss of immersion would have a negative impact:

'I think when you're starting with something new, you've got the distraction of the vehicle which it's presented in and so you're not immersed because you're also worrying about, 'Am I doing it right? Is it the right way round?' And the immersion's all about being able to focus on the thing that you're interested in and that's because you're not having to worry about other things [...] I think people would feel more immersed if they did it more often because they would start to be familiar with it and they then wouldn't notice.' (HC4)

Interesting in the interviewees' comments was the different ways in which they conceptualised the sense of immersion – whether it comes from having all visual attention focused on the slide, or from the physical connection with the microscope, or simply from familiarity with the technology of the conventional microscope, so that the user does not have to think about what they are doing. However, five of the interviewees did not present this as a concern.

Perceived impact on efficiency

When asked whether they thought that the introduction of digital pathology would have an impact on efficiency, three interviewees said that they felt the impact on efficiency would be a positive one, although all three identified different sources for this improvement. One interviewee talked about efficiency in terms of not having to wait for referred cases (assuming that the referring site also used digital slides), while another talked of removing the physical transfer of slides. Another focused on the benefits of looking at a screen as opposed to looking down a microscope:

'So [the microscope] is a […] kind of tunnelled vision or tunnelled thinking process which could be easily distracted. Whereas if […] you're looking down the screen, even if for a moment I want to look out of the window you can still come back and that is still there in a wider view staring at you.' (HC6)

Willingness to move to digital

There was huge variation in interviewees' enthusiasm for digital pathology in relation to routine diagnostic work. While most could be described as open to the idea, one interviewee rejected the idea strongly:

'We haven't got enough money to do what we're doing now. And the investment that you'd need to really make digital pathology taken on as for routine I think is unjustifiable.' (HC3)

At the other extreme, we had one interviewee who was very keen to move to using digital pathology as soon as possible:

'These [gesturing to microscope] can all go in the scrap heap. I just think it liberates us so much as well. […] I think it's just going to be revolutionary actually, quite frankly, I really do.' (HC9)

One interviewee acknowledged that there would be a learning curve but did not see this as a problem:

'So if you are going to give a diagnosis to a patient, are you ready to sign out on the screen? I said yeah, maybe if I start using it, it's only a question of getting used to it, I can't see any great difficulty there.' (HC6)

Generally, the impact on safety was not a concern amongst the pathologists that we interviewed. Four of the interviewees felt that there would not be a negative impact on safety, with three of the four saying that safety might be improved by reducing the risk of mixing up slides, while one interviewee was unsure what the impact on safety would be. There was little concern regarding image quality:

'An image on a computer if it's high quality is no more or less real than you know a section that's been stained and has light shone through it really, […] it's just a representation of what was happening in life anyway, isn't it? Not really live tissue, just a thin slice of it that's been stained.' (HC5)

Concerns regarding safety related to lack of familiarity with the technology:

'There's the patient safety from getting it wrong using a system you're not used to or haven't been trained in and therefore there could be some drawbacks which we're just not used to. […] Stain deposits, dodgy staining, […] how does dodgy staining look? 'cos we're so used to looking down the microscope. And you've been looking at it for like 10 years.' (HC2)

Three of the interviewees said that they would want the opportunity to try it out for themselves:

'I'd just have to had seen quite a number [of digital slides] I think, probably in parallel [with glass slides], or alternate, or something like that. Just to see if you would have made the same diagnosis. And er, yeah, just to get a feeling for it, 'cos it's a different system. The trouble is, with the glass, you've been doing it for so long you're used to all the flaws and you compensate for all the flaws. Whereas with the computer system, you haven't used it, you might not know all the niggles and glitches and flaws.' (HC2)

One interviewee said that he would like to see data regarding the time it takes to become familiar with the system:

'I think the kind of data to be after would be, that would be worth getting, would be to be able to demonstrate to people who have not used it very much that there is a learning curve and that people can approach the diagnostic speeds that they are used to.' (HC4)

One interviewee was more concerned about the processes surrounding the technology, in terms of the training and support that would be provided. This seemed to be influenced by their experience of other IT systems within the hospital:

'I just think you've got to have that IT support to sort it out straight away. [...] and you need to pay people proper money to support it. And our IT guy's just had a 20% pay cut because he's been down-banded, you know? [...] That's going to impact on how well your system works. [...] That's the other thing, we get [lab system] or whatever and you start learning new things about it years later. 'Oh I didn't know you could do that.' Quite earth shattering things that can drastically reduce your amount of time. So, you know, people need to be trained properly in it...' (HC5)

Robustness was an issue mentioned by three of the interviewees, two of whom explicitly linked this to their experience of hospital IT systems:

'I think that's probably my main concern. […] They're very good at being optimistic but the delivery's always a major disappointment.' (HC7)

Discussion

Interviewees were able to identify a range of potential benefits of digital pathology in the context of diagnostic work, with a particular emphasis on easier access to slides and the efficiency gains this could bring. Amongst the barriers to use of digital pathology, a key concern was lack of familiarity, not only in terms of becoming familiar with the technology but learning how to adjust their diagnostic skills to this new medium. This reflects similar findings from radiology, where the reading of images takes place in light of readers' knowledge of the principles of the production of the image as well as specific local practices, allowing them to distinguish between artifacts of the screening process and those that have diagnostic significance [23]. Findings from the experimental studies we have undertaken suggest that pathologists need to relearn what they should expect to see at each magnification level, as it may be that it is necessary to view digital slides at a higher level of magnification than would typically be required for viewing glass slides on a conventional microscope [18].

Interestingly, interviewees did not express concern about diagnostic performance, in contrast to previous studies [9]. They also did not have concerns about the speed of digital slides, despite this being highlighted as an issue in previous studies [15-17], instead focusing on the potential efficiency gains of a digital workflow. This may be due to the interviewees' lack of familiarity with digital slides, their responses suggesting a belief that, with experience, the speed and accuracy of their diagnoses would increase. On this basis, we would suggest that the majority of the interviewees in this study were fairly positive about the idea of using digital pathology for diagnostic work but were rightly cautious about its impact on their work, wanting more opportunity to try the technology themselves, alongside the microscope so that they could judge for themselves the accuracy of their diagnoses, and aware that it is not just an issue of learning to use the technology but potentially has implications for how they make diagnoses.

Where negative attitudes to digital pathology were expressed, was on the basis that glass slides work adequately and so it may be difficult to justify the investment of money and time (in learning to use the technology) that a move to digital pathology for routine diagnostic work would require. Greater evidence of cost savings that digital pathology can provide may go some way to overcoming such concerns. However, challenges often arise in adoption of technology when the users of the technology are not those who will benefit most. The key benefits of digital pathology are currently predicted to be in workflow efficiency, which benefits the organisation, and potentially the patient, rather than providing any efficiency benefit to the individual. However, developments such as the integration of decision support could provide benefit to the user [9,10]. Perception of benefits is likely to vary according to organisational factors; our institution is a tertiary centre providing specialist expertise, and so pathologists do not often need to seek advice from pathologists at other institutions where digital pathology would provide a benefit.

Another challenge is that, while previous studies have compared adoption of digital images in radiology and pathology [9], in radiology the images begin as digital data, whereas for digital pathology the need to create the glass slide remains, so that digital pathology introduces an extra step into the process [10]. However, in the future it may be possible to produce digital slides without the need to first produce a glass slide, for example by using spectral domain optical coherence tomography (SD-OCT) to scan tissue blocks [24].

Implications for practice

The findings suggest, as with any health IT system, adequate training and support need to be in place for digital pathology to be effectively integrated into diagnostic work. Before transitioning to a totally digital workflow, pathologists may benefit from the opportunity to review glass slides alongside digital images as well as learn to use the technology, that can help to adjust their diagnostic skills and gain confidence in their ability to make a correct diagnosis with a digital slide. There is a need for further research on the learning curve associated with digital pathology, to reassure pathologists about the time investment required to work in this new way. This is work that we have begun to undertake [25].

Strengths and limitations

This is a small scale study, conducted within a single institution, so we cannot judge to what extent the findings reflect the attitude of pathologists in general. However, our participants were from a range of subspecialties and had varying levels of experience of digital pathology.

A strength of our study is that it provides detailed qualitative data on the barriers to integrating digital pathology into diagnostic work, with larger number of participants compared to previous studies, and provides guidance for those who seek to implement such systems.

Conclusion

Interviewees were able to identify a range of potential benefits of digital pathology in the context of diagnostic work, with a particular emphasis on easier access to slides and the efficiency gains this could bring. They were predominantly positive about the idea of using digital pathology for diagnostic work but rightly cautious about its impact on their work, aware that it is not just an issue of learning to use the technology but potentially has implications for how they make diagnoses.

Acknowledgements

We would like to thank those pathologists who have allowed us to observe their work and who have taken part in our evaluations of the LVM. This report is independent research commissioned by the National Institute for Health Research under NEAT. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. The authors acknowledge the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

References

[1] Georgiou A, Westbrook J, Braithwaite J. An empirically-derived approach for investigating Health Information Technology: the Elementally Entangled Organisational Communication (EEOC) framework. BMC Medical Informatics and Decision Making 2012; 12, 68.

[2] Hanseth O, Lundberg N. Designing Work Oriented Infrastructures. Computer Supported Cooperative Work (CSCW) 2001; 10, 347-372.

[3] Karasti H. Bridging Work Practice and System Design: Integrating Systemic Analysis, Appreciative Intervention and Practitioner Participation. Computer Supported Cooperative Work (CSCW) 2001; 10, 211-246.

[4] Johnson R, O'Hara K, Sellen A, Cousins C, Criminisi A. Exploring the Potential for Touchless Interaction in Image-Guided Interventional Radiology. CHI 2011. Vancouver, BC: ACM, 2011.

[5] Mentis H, O'Hara K, Sellen A, Trivedi R. Interaction Proxemics and Image use in Neurosurgery. CHI 2012. Austin, Texas: ACM, 2012.

[6] Gilbertson J, Ho J, Anthony L, Jukic D, Yagi Y, Parwani A. Primary histologic diagnosis using automated whole slide imaging: a validation study. BMC Clinical Pathology 2006; 6, 4.

[7] Della Mea V, Demichelis F, Viel F, Dalla Palma P, Beltrami CA. User attitudes in analyzing digital slides in a quality control test bed: A preliminary study. Computer Methods and Programs in Biomedicine 2006; 82, 177-186.

[8] Nakhleh RE. Patient Safety and Error Reduction in Surgical Pathology. Arch Pathol Lab Med 2008; 132, 181-185.

[9] Patterson ES. Barriers and facilitators to adoption of soft copy interpretation from the user perspective: Lessons learned from filmless radiology for slideless pathology. Journal of Pathology Informatics 2011; 2, 1.

[10] Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC, Collins LC, Colgan TJ. Review of the current state of whole slide imaging in pathology. J Pathol Inform 2011; 2, 36.

[11] Blake CA, Lavoie HA, Millette CF. Teaching medical histology at the University of South Carolina School of Medicine: Transition to virtual slides and virtual microscopes. The Anatomical Record Part B: The New Anatomist 2003; 275B, 196-206.

[12] Kumar RK, Velan GM, Korell SO, Kandara M, Dee FR, Wakefield D. Virtual microscopy for learning and assessment in pathology. The Journal of Pathology 2004; 204, 613-618.

[13] Dennis T, Start RD, Cross SS. The use of digital imaging, video conferencing, and telepathology in histopathology: a national survey. J Clin Pathol 2005; 58, 254-258.

[14] Furness P. A randomized controlled trial of the diagnostic accuracy of internet-based telepathology compared with conventional microscopy. Histopathology 2007; 50, 266-273.

[15]Onega T, Weaver D, Geller B, Oster N, Tosteson AA, Carney P, Nelson H, Allison K, O'Malley F, Schnitt S, Elmore J. Digitized Whole Slides for Breast Pathology Interpretation: Current Practices and Perceptions. Journal of Digital Imaging 2014; 27, 642-648.

[16] Treanor D, Quirke P. The Virtual Slide and Conventional Microscope - a Direct Comparison of Their Diagnostic Efficiency. The Journal of Pathology 2007; 213, 7a.

[17] Velez N, Jukic D, Ho J. Evaluation of 2 whole-slide imaging applications in dermatopathology. Human Pathology 2008; 39, 1341-1349.

[18]Randell R, Ruddle RA, Mello-Thoms C, Thomas RG, Quirke P, Treanor D. Virtual reality microscope versus conventional microscope regarding time to diagnosis: an experimental study. Histopathology 2013; 62, 351-358.

[19] Randell R, Ruddle RA, Quirke P, Thomas RG, Treanor D. Working at the microscope: analysis of the activities involved in diagnostic pathology Histopathology 2012; 60, 504-510.

[20] Randell R, Ruddle RA, Thomas R, Treanor D. Diagnosis at the microscope: a workplace study of histopathology. Cognition, Technology & Work 2012; 14, 319-335.

[21]Randell R, Ruddle RA, Thomas RG, Mello-Thoms C, Treanor D. Diagnosis of major cancer resection specimens with virtual slides: impact of a novel digital pathology workstation. Human Pathology 2014; 45, 2101-2106.

[22] Pope C, Ziebland S, Mays N. Analysing qualitative data. In: Pope C, Mays N, eds. Qualitative Research in Health Care. Oxford: Blackwell Publishing/BMJ Books, 2006, p. 63-81.

[23] Jirotka M, Procter R, Hartswood M, Slack R, Simpson A, Coopmans C, Hinds C, Voss A. Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study. Computer Supported Cooperative Work (CSCW) 2005; 14, 369-398.

[24] Fine JL, Kagemann L, Wollstein G, Ishikawa H, Schuman JS. Direct scanning of pathology specimens using spectral domain optical coherence tomography: a pilot study. Ophthalmic Surg Lasers Imaging 2010; 41 Suppl, S58-64.

[25] Blizzard R, Ruddle R, Thomas R, Randell R, Treanor D. A longitudinal evaluation of histopathologists' performance using the Leeds Virtual Microscope. 26th European Congress of Pathology. London, 2014.

Address for correspondence

Dr Rebecca Randell, School of Healthcare, Baines Wing, University of Leeds, Leeds LS2 9UT, UK. r.randell@leeds.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-448

Big Data Clinical Research: Validity, Ethics, and Regulation

E. Andrew Balas[a], Marlo Vernon[a], Farah Magrabi [b], Lynne Thomas Gordon[c], Joanne Sexton[a]

[a] *Georgia Regents University, Augusta, GA, USA*
[b] *Australian Institute of Health Innovation, Macquarie University, NSW, Australia*
[c] *American Health Information Management Association, Chicago, IL, USA*

Abstract

Electronic Health Records (EHR) promise improvement for patient care and also offer great value for biomedical research including clinical, public health, and health services research. Unfortunately, the full potential of EHR big data research has remained largely unrealized.

The purpose of this study was to identify rate limiting factors, and develop recommendations to better balance unrestricted extramural EHR access with legitimate safeguarding of EHR data in retrospective research. By exploring primary, secondary, and tertiary sources, this review identifies external constraints and provides a comparative analysis of social influencers in retrospective EHR-based research.

Results indicate that EHRs have the advantage of reflecting the reality of patient care but also show a frequency of between 4.3-86% of incomplete and inaccurate data in various fields. The rapid spread of alternative analytics for health data challenges traditional interpretations of confidentiality protections. A confusing multiplicity of controls creates barriers to big data EHR research.

More research on the use of EHR big data is likely to improve accuracy and validity. Information governance and research approval processes should be simplified. Comprehensive regulatory policies that do not exclusively cover health care entities, are needed. Finally, new computing safeguards are needed to address public concerns, like research access only to aggregate data and not to individually identifiable information.

Keywords:

Electronic Health Records; Clinical Research; Public Health; Health Services Research.

Introduction

The vast amount of clinical data accumulating in Electronic Health Records (EHRs), or big data EHRs, represents an unprecedented opportunity to discover unrecognized risk factors, study the epidemiology of diseases, calculate life expectancy, distinguish best practices from superior outcomes and recognize opportunities for better health care.

The review of patient charts has been the cornerstone of clinical research for centuries. Historically, many landmark discoveries have originated from analyses of retrospective data. The relationship between smoking and lung cancer was first discovered by Müller in 1941 based on an analysis of patient records [1]. From these simple beginnings, recognition of smoking as a health hazard continued to evolve, ultimately becoming one the greatest public health achievements of the 20th century [2]. More contemporary examples include the relationship between thalidomide and birth defects [3], cancer epidemiology and pathophysiology, and vaccine development [4-6].

Facilitating biomedical research is one of the most important, but unrealized promises of introducing EHRs [7]. Researchers interested in conducting needs analysis, process, and outcome evaluations utilizing EHRs often run into seemingly insurmountable barriers.

Extramural access to big data represents a particular challenge (e.g., researcher of institution A trying to study big data of institution B). An illustrative case from a recent personal communication:

> *A professor of one of the world's top ranked universities wanted to get a large EHR data set for research from a leading hospital nearby. The request for an anonymized data set and the study plan was approved by the institutional IRB. In spite of regulatory compliance, ethics clearance, and stellar personal scientific track record, the professor was unable to obtain the EHR data over a period of 18 months and finally gave up.*

With the advancement of computer hardware and software, access to and analysis of clinical data is no longer a primarily technical issue. Today, the principal obstacles to EHR use in research are essentially social, ethical and regulatory.

Significant gaps exist for researchers in requesting, accessing, analyzing, and applying EHR data [8]. The Institute of Medicine reports that disappointment in the lack of EHR improvements has tempered enthusiasm for continuing research efforts [9]. Patients/consumers are increasingly participating in their own care and in care decisions, and most report being open to sharing their data for research [10, 11]. Their priorities regarding use and re-use of data will need to be taken into consideration.

Ideally, EHR-based research should meet the simultaneous but somewhat conflicting requirements of both unrestricted extramural access to the EHR by meritorious, innovative biomedical researchers with minimal administrative requirements and guaranteed zero access to EHRs by unauthorized, unnecessary or potentially harmful users of health data. Obviously, unnecessary and unjustified limitations on the access to EHR big data also represents a serious ethics violation in terms of denied care, lost public health improvement, and unrealized research discoveries benefiting patients.

The purpose of this study is to identify rate limiting social factors and develop recommendations to facilitate research on EHR big data. This study focuses on three dimensions: validity, ethical considerations, and security risks of EHR data use in biomedical research, focusing on the US context.

Methods

The project explores relevant studies and methods originating in biomedicine, health informatics, historical research, bioethics, health administration, computer science, public health and other fields. Eligibility criteria: EHR-based original research project of at least 1000 records, or thematic exploration of EHR big data management in research. Peer reviewed literature and policy documents were explored along the following hierarchy:

1. Primary sources (data): original research publications in the peer-reviewed scientific literature on EHR projects, national and international statistical databases, national surveys, and historical documents illustrating important aspects of EHR use in research.
2. Secondary sources (management): thematic scholarly explorations in the scientific literature, critiques, pertinent scholarly books, government documents, and statements of national and international organizations on social actors in big data research.
3. Tertiary sources (pointers): newspaper articles, speeches, videos, encyclopedias, web pages and other popular publications that help to identify primary and secondary sources of information.

In processing and synthesizing the information, tertiary sources were used solely to identify primary and secondary sources of information, which serve as the backbone; meanwhile, secondary sources were used to elaborate and enhance the ideas and themes of the primary sources.

PubMed and Web of Science were utilized for full text searches. Search terms included: electronic health records (EHR), electronic medical record (EMR), retrospective health research, ethics of EHR data use, confidentiality of EHR data and HIPAA. We then investigated similar themes and ethical dimensions presented in the literature. A PRISMA framework was applied to the larger systematic review resulting from this analysis [12]. The comparative and discerning analyses generated the list of factors that create external and internal influences on big data EHR use.

Results

Patient data from large EHR databases are increasingly available from multiple sources and often for a price. Efforts to evaluate the availability of these data from a practical and ethical standpoint is ongoing (see Table 1).

Validity - Issues of data integrity

Data integrity is defined as the validity, accuracy, reliability, timeliness, and consistency of the data. It remains the first question of recorded EHR data use in biomedical research. Retrospective analyses need to consider the limitations and appropriate use of data, including potential risks of inaccuracies [8]. Retrospective data are collected in variable circumstances, recorded with inconsistent data definitions, missing data, and without standardized testing. Table 2 lists the frequency of some of the reported deficiencies in EHR-based research.

On the other hand, this patient care data represents the reality of actual practice, as opposed to results of sterile research protocols. In many cases, real data can fill gaps in current evidence and provide evidence in areas where clinical trials will never be carried out. Illustrative and appropriate research uses of retrospective clinical data include exploration of risk factors, cost-effectiveness of care, selection of best practices, and the epidemiology of diseases and health conditions.

Table 1 - Selected EHR data aggregators

Source	Type of Data
CMS	Medicare and Medicaid
Blue Health Intelligence	Claims data on 210 million individuals, available longitudinally
Aetna – Accountable Care Solutions	Claims data on Aetna subscribers
Validic	Commercial firm, data aggregator for physicians and health systems
Kaiser Permanente Health Connect (Northern CA)	9.1 million patients Subscriber health claims data
Massachusetts Health Quality Partners	Intramural, work with Department of Public Health and others
OCHIN	Members of 70 health system across 19 states
IMS® Disease Analyzer	EHR are contributed by a representative panel of more than 2.500 physicians in Germany
Humana Health Care – Anvita Health	11.2 million members health data
Cerner Health Facts	Since 2000, EHR collected from 480 contributing facilities throughout USA
Vestrum	EHR data from private physicians
MS HealthVault	Personal health information of "far more" than the tens of thousands of users
Express Scripts and CVS Pharmacies	Sale of prescription information which is “match-backed” by third parties, and linked to website databases

Table 2. – Frequency of deficiencies in EHR-based research

Source	Estimate	Reference
Incompleteness	24%	[11, 13, 14]
	86%	[15]
	65%	[16]
	86%	[17]
CPOE Errors	51.4-91.5	[18]
Inaccuracies, errors	4.3 %	[19]
Inconsistencies	variable	[11, 13, 14]

EHR-based analyses provide the opportunity to evaluate individual outcomes and compare results to larger populations [20]. Reduction in costs comes with streamlined processes and improved practice efficiency. The National Patient-Centered Clinical Research Network hopes to provide unparalleled access by producing national EHR data sets available to researchers [21].

Research on rare diseases is difficult because of small sample sizes, which may be geographically distant from each other. Expanded EHR networks would enable easier access for both patient and provider, and in turn increase opportunities for research [22, 23].

The gold standard of evidence based medicine, multi-center randomized controlled clinical trials (RCT), are enormously expensive and time consuming, particularly when the sample size is very large to support high power analyses [24]. Due to

special resources and arrangements, RCTs often represent centrally-controlled practices that are not representative of current practices; and also, may never be fully replicated in general use. Furthermore, prospective randomized studies can evaluate only beneficial interventions, as opposed to allocation of patients to a harmful intervention, such as smoking, which would be unethical.

In the assessment of biological phenomena and therapeutic potential, the multicenter RCT cannot be replaced by anything less methodologically rigorous. However, due to the very high rate of negative clinical trial results, the use of retrospective data from EHR is recommended for more effective filtering prior to the evaluation of safety and efficacy of new treatments in prospective studies [25].

The use of large databases requires special statistical techniques as enormous sample sizes can overpower results (i.e., practically meaningless minutia can appear statistically significant). Additionally, appropriate scrutiny of data quality and accuracy variability is needed with reasonable logistical and statistical checking prior to analyses.

Ethics: Issues of Privacy, Beneficence, and Non-maleficence

Ethics concerns in big data biomedical research are twofold: wrongdoing in research, and ineffective administration of human subject reviews. In this study, we referred to the DuBois taxonomy of misbehavior in medical research [26]. Out of 15 fundamental kinds of wrongdoings in medical research, two stand out as particularly relevant to big data research (Table 3).

Table 3. Taxonomy of Misbehavior in Medical Research

Application Area	Examples
Violation of privacy or confidentiality	• wrongful disclosure • wrongfully obtaining information • wrongful use of health information • failure to safeguard health information
Failure of informed consent	• no need for such consent regarding archived data • advance blanket research approval forms to facilitate future use

Privacy is the basic human right of limited access by others to aspects of their person, including thoughts, identifying information, and even information in bodily tissues and fluids [27]. Patients are increasingly playing an active role in their care and are often unaware that their health information, de-identified or not, may be used for clinical or community health research in the future. This directly impacts the autonomy of the patient, even if they would have given consent.

Confidentiality is the mandate to protect information that an individual has disclosed in a relationship of trust [27]. In certain circumstances, personal information may be analyzed without consent when the benefits to society outweigh the individual's interest in keeping the information confidential [28]. Typically, big data researchers are not involved in data collection and, therefore, confidentiality and security of protected data become the foremost concerns.

Extramural research is much harder to conduct than collaborating with an intramural colleague or being an inside user of local databases. Institutional Review Boards are tasked to protect human subjects in research. Unfortunately, variable interpretations and a lack of coordination among multi-site IRBs creates a challenging health research environment [7]. This has been recognized by the NIH as an area in need of improvement. The recently published draft policy for public comment noted "there is no evidence that multiple IRB reviews enhance protections for human subjects" [29].

Regulation: Data Security and Alternative Analytics

In the US, the Health Insurance Portability and Accountability Act (HIPAA) and subsequent regulations direct covered entities (e.g., a healthcare institutions) in the protection of individually identifiable protected health information in research. Violations of the Privacy Rule can become the basis for both civil and criminal penalties, including fines and possible time in jail.

A particularly controversial part of the HIPAA provisions is use of de-identified data in research. For facilitated research access, the Privacy Rule requires de-identification (i.e., removal of 18 identifiers including names, social security numbers, telephone numbers, and others). However, de-identification creates obstacles to many research projects, particularly cause-effect and time series studies.

Table 4. Examples of alternative analytics

Source	Type of Data	Alternative Use
28% of US hospitals	Patient wealth screening	Grateful Patient Program
Target	Consumer data	Use of shopping pattern identifies marketing strategies, including based on health behaviors: pregnancy, diabetes
Garmin Connect	Athletic performance data	4 billion miles of performance information
CRM Healthgrades	Aggregate health data	Sells patient lists based on diagnosis, evaluates hospital patient data for non-compliance and QC
Carolinas HealthCare	Consumer data on 2 million people	Identify high-risk patients. Data aggregated through public records, store loyalty program transactions, and credit card purchases.
LexisNexis	Medicaid recipients and consumer data publicly available (vehicle registration, property records, etc.)	Identify Medicaid Fraud and Abuse

Without an authorization, a covered entity can use and disclose protected health information for treatment, payment and health care operations (TPO). Recently, the American Medical Informatics Association started advocating for the inclusion of observational or non-interventional data research as an appropriate operational use of protected health information [30, 31]. Ultimately, EHR big data should be available not just in the present, but also for improved future decision-making.

The rapid spread of alternative analytics for health records also challenges traditional interpretations of covered entity and confidentiality protections. For example, the recently announced Qualcomm Tricorder XPRIZE is a $10 million global competition to accurately diagnose a set of diseases independent of a healthcare professional or facility. Corporations have discovered algorithms in shopping and browsing patterns, utilizing shopping cards and credit card transactions, to identify health diagnosis or needs – without the need to access a person's medical record (Table 4). While this is skillful marketing and consumer targeting, it appears that little thought has been given to the ethics of such analysis.

Data matching with publicly available records is also becoming possible [32]. While the use of this same consumer data when legitimately coupled with a person's health record may have genuine positive outcomes, such as warning that purchased food may interact poorly with current medications or reminders to refill prescriptions, the negative consequences are not far-fetched.

The "Grateful Patient Program" model is gaining more support across the country as a way to increase donations to both non-profit and for-profit hospitals [33]. Offices of giving or communication match EHR data with income/wealth data to identify prospective donor patients and families, and then send targeted information, or even organize special visits and contacts when the patients are admitted [34]. Currently, alternative analytics has far fewer regulatory obstacles than big data biomedical research.

In addition to the ever increasing flow of health information with data stored on hard drives and cloud computing, human tissues and cells also challenge the current system of protections as they are also carriers of personally identifiable health information.

The American Health Information Management Association recommends comprehensive information governance to ensure accuracy, reliability, integrity, timeliness, accessibility, and security of data and information impacting patient care, research studies and public policy. Health care data and information must be governed to meet these imperatives.

Discussion

Big data EHRs have proved to be not only an irreplaceable source of clinical, public health and health services research but also an epicenter of confusing expectations and restrictions. The balance between access to EHRs, validity concerns, ethics safeguards and regulatory protections remains elusive.

To unlock the full potential of big data EHR research, a series of actions are needed:

1) More research on the use of EHR big data is not only a matter of new scientific knowledge but also immediate public interest as more use is likely to discover more errors and stimulate corrective efforts [35]. More EHR big data research is likely to improve accuracy and validity through improved error detections and control mechanisms.

2) The confusing multiplicity of controls creates hindrances in big data EHR research (e.g., IRB for human subject protection, institutional privacy officers for regulatory compliance, multiple IRBs for inter-institutional collaborations). Therefore, information governance and the research approval processes should be integrated and streamlined to be a one-stop research approval process.

3) Considering the rapidly expanding array of health databases outside the health care system, and alternative analytics, there is the risk that academic research and also patient care by licensed clinicians will be outpaced and major ethical and security concerns will be unaddressed. Comprehensive policies are needed for secondary use of all electronic health data, not just those in currently covered health care entities.

4) The many rate-limiting privacy and information security requirements call for new computing safeguards to address public concerns (e.g., creation of access only to aggregate data but not to individually identifiable information, tools for matching big data from diverse sources without revealing individual data).

Trust in the protection and utilization of health data is essential to continued public support. The public must see and believe that their data security is taken seriously and used responsibly. It is reasonable to expect that the larger availability of EHRs and the opportunities to match data from multiple databases should lead to an accelerated rate of valuable research discoveries.

References

[1] Müller F. Tabakmißbrauch und Lungencarcinom. Z Krebs-forsch. 1940 1940/01/01;49(1):57-85. German.

[2] Centers for Disease Control and Prevention. Ten great public health achievements--United States, 1900-1999. MMWR Morbidity and mortality weekly report. 1999;48(12):241.

[3] Lenz W, Knapp K. Thalidomide embryopathy. Archives of environmental health. 1962 Aug;5:100-5. Epub 1962/08/01. eng.

[4] Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. Nature medicine. 2011;17(3):297-303.

[5] DeVita Jr VT, Rosenberg SA. Two hundred years of cancer research. New England Journal of Medicine. 2012;366(23):2207-14.

[6] zur Hausen H. Papillomaviruses in the causation of human cancers—a brief historical account. Virology. 2009;384(2):260-5.

[7] AHRQ. A Robust Health Data Infrastructure. Rockville, MD: Agency for Healthcare Resarch and Quality,, Director HI; 2014 April 2014. Report No.: 14-0041-EF,.

[8] Roth CP, Lim Y-W, Pevnick JM, Asch SM, McGlynn EA. The Challenge of Measuring Quality of Care From the Electronic Health Record. American Journal of Medical Quality. 2009 September 1, 2009;24(5):385-94.

[9] Institute of Medicine. Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary. Washington DC: National Academy of Sciences; 2010.

[10] Alston C, Paget L, Halvorson G, Novelli B, Guest J, McCabe P, et al. Communicating with patients on health care evidence. Institute of Medicine, Washington [DC]. 2012.

[11] Walker J, Meltsner M, Delbanco T. US experience with doctors and patients sharing clinical notes. BMJ. 2015;350(g7785).

[12] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare

interventions: explanation and elaboration. Bmj. 2009;339:b2700.

[13] Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. AMIA summits on translational science proceedings. 2010;2010:1.

[14] Denham CR, Classen DC, Swenson SJ, Henderson MJ, Zeltner T, Bates DW. Safe use of electronic health records and health information technology systems: trust but verify. Journal of patient safety. 2013 Dec;9(4):177-89. Epub 2013/11/22. eng.

[15] Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. Bmj. 2003 May 17;326(7398):1070.

[16] Kopcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T, et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. BMC medical informatics and decision making. 2013;13:37.

[17] McGinnis KA, Skanderson M, Levin FL, Brandt C, Erdos J, Justice AC. Comparison of two VA laboratory data repositories indicates that missing data vary despite originating from the same source. Medical care. 2009 Jan;47(1):121-4.

[18] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. Jama. 2005 Mar 9;293(10):1197-203.

[19] Weiss J, Kumata J, Galar A, Turkish L, editors. Postoperative Eye Drop Documentation Omissions With EHRs After Resident Cataract Surgery: An Underrecognized Source of Error. American Academy of Opthalmology Annual Meeting; 2014 October 14, 2014; Chicago, IL.

[20] Schroeder EB, Goodrich GK, Newton KM, Schmittdiel JA, Raebel MA. Implications of different laboratory-based incident diabetic kidney disease definitions on comparative effectiveness studies. Journal of comparative effectiveness research. 2014 Jul;3(4):359-69.

[21] Terry K. Giant EHR-based network will compare treatments. The National Patient-Centered Clinical Research Institute's program may allow physicians to compare how treatments work. Medical economics. 2014 Jun 25;91(12):38-41.

[22] Bowles KH, Potashnik S, Ratcliffe SJ, Rosenberg M, Shih NW, Topaz M, et al. Conducting Research Using the Electronic Health Record Across Multi-Hospital Systems: Semantic Harmonization Implications for Administrators. The Journal of nursing administration. 2013 Jun;43(6):355-60.

[23] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association : JAMIA. 2007 Jan-Feb;14(1):1-9. Epub 2006/11/02. eng.

[24] Concato J. When to randomize, or 'Evidence-based medicine needs Medicine-based evidence'. Pharmacoepidemiology and Drug Safety. 2012;21:6-12.

[25] Olsen L, McGinnis JM. Redesigning the Clinical Effectiveness Research Paradigm:: Innovation and Practice-Based Approaches: Workshop Summary: National Academies Press; 2010.

[26] DuBois JM, Kraus E, Vasher M. The development of a taxonomy of wrongdoing in medical practice and research. American journal of preventive medicine. 2012 Jan;42(1):89-98.

[27] Plaza J, Fischbach R. Current Issues in Research Ethics : Privacy and Confidentiality [Online Learning]. New York, NY: Columbia University, Columbia Center for New Media Teaching and Learning,; [cited 2014 December 15]. Available from: http://ccnmtl.columbia.edu/projects/cire/pac/introduction/index.html.

[28] Lowrance WW. Learning from experience: privacy and the secondary use of data in health research. The journal of biolaw & business. 2003;6(4):30-60.

[29] National Institutes of Health. Request for Comments on the Draft NIH Policy on the Use of a Single Institutional Review Board for Multi-Site Research 2014 [cited 2014 December 8]. Available from: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-026.html.

[30] Middleton B, Martin R, AMIA. HIPAA Changes Needed to Permit Researcher Access to Patient Records, Letter to Rep. Fred Upton re: Request for Comments on Energy and Commerce Digital Health White Paper, 2014 [updated 10/15/2014; cited 2014 December 12]. Available from: http://www.amia.org/sites/amia.org/files/AMIA-21st-Century-Cures-Comments-to-House-EC%20-%202014-10-15.pdf.

[31] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association. 2007 1//;14(1):1-9.

[32] Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing Health Data for Better Outcomes on PatientsLikeMe. Journal of Medical Internet Research. 2010 Apr-Jun;12(2):e19.

[33] Stewart R, Wolfe L, Flynn J, Carrese J, Wright SM. Success in Grateful Patient Philanthropy: Insights from Experienced Physicians. The American Journal of Medicine.124(12):1180-5.

[34] Elj T. Grateful patient programs: current trends, strategies and tactics. AHP journal/Association for Healthcare Philanthropy. 2006:8-9, 11, 3 passim.

[35] Miriovsky BJ, Shulman LN, Abernethy AP. Importance of Health Information Technology, Electronic Health Records, and Continuously Aggregating Data to Comparative Effectiveness Research and Learning Health Care. Journal of Clinical Oncology. 2012 December 1, 2012;30(34):4243-8.

Address for correspondence

E. Andrew Balas, MD, PhD, Georgia Regents University, 987 St. Sebastian Way, EC 3423, Augusta, GA 30912, Office: 706-721-2621, Fax: 706-721-7312, Email: andrew.balas@gru.edu.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-453

Streaming Physiological Data: General Public Perceptions of Secondary Use and Application to Research in Neonatal Intensive Care

Carolyn McGregor[a], Jennifer Heath[b], Yvonne Choi[c]

[a] *Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Canada*
[b] *Business Analysis & Learning Analytics, University of Wollongong, Wollongong, Australia*
[c] *Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa, Canada*

Abstract

High speed physiological data represents one of the most untapped resources in healthcare today and is a form of Big Data. Physiological data is captured and displayed on a wide range of devices in healthcare environments. Frequently this data is transitory and lost once initially displayed. Researchers wish to store and analyze these datasets, however, there is little evidence of any engagement with citizens regarding their perceptions of physiological data capture for secondary use. This paper presents the findings of a self-administered household survey (n=165, response rate = 34%) that investigated Australian and Canadian citizens' perceptions of such physiological data capture and re-use. Results indicate general public support for the secondary use of physiological streaming data. Discussion considers the potential application of such data in neonatal intensive care contexts in relation to our Artemis research. Consideration of the perceptions of secondary use of the streaming data as early as possible will assist in building appropriate use models, with a focus on parents in the neonatal context.

Keywords

Patient data privacy; Data collection; Medical device

Introduction

In recent research, there has been an increased interest in the analysis of physiological data, particularly in real-time. Many critical care and neurological monitoring applications capture physiological data. Examples include electrocardiogram (ECG), electroencephalogram (EEG) and pulse oximetry data [1].

High speed physiological data represents one of the most untapped resources in healthcare today and is a form of Big Data. In neonatal intensive care for example, a premature newborn infant's heart beats approximately 7000 times an hour and yet traditional charting on paper, or within an electronic health record (EHR), includes one number per hour of an indicative heart rate for that hour. The heuristics employed to determine the number to write are as much qualitative as quantitative and part of the function is to express overall stability or instability hour to hour. The potential for the use of high speed physiological data for earlier and potentially more reliable pathophysiological indicators have been presented for late onset neonatal sepsis [2], pneumothorax [3], intraventricular haemorrhage [4, 5] and periventricular leukomalacia [6]. Opportunities abound for the exploration of new pathophysiological indicators for many other conditions, but these are yet to be explored due to the absence of a collection of physiological data for patients developing such conditions. Analytics on high frequency physiological data, from both the perspective of retrospective knowledge discovery and real-time monitoring has the potential to be equally disruptive for healthcare as genomics research.

Within the Neonatal Intensive Care Unit (NICU), a variety of medical devices monitor the infant's vital organs while other equipment assist with breathing, maintains appropriate body temperature and provides necessary drugs and nutrients. Many of these devices continuously create physiological data second-by-second. Although many health care settings are in the process of transitioning from the use of paper to electronic for charting purposes, this does not include new approaches for new analytics derived from this sensor data. This leads health care professionals to rely on sharing clinical information in a qualitative manner [7]. Comparable to the concept of business intelligence and analytics, which stems from the prompt interpretation of large volumes of data for actionable information, there is a growing urgency for the health care sector to similarly adopt a notion of "health care intelligence" in (near) real-time [1]. There are limitations regarding the use of analytics in health care due to the time it takes to deliver predictions to healthcare providers and enable action [8].

The value of Big Data comes from the ability to make "connections between pieces of data, about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself" [9]. An example of a Big Data platform that includes the use of physiological streaming data is Artemis. Artemis supports online health analytics that allows for concurrent multi-patient, multi-diagnosis and multi-stream temporal analysis of complex, high-frequency physiological data streams in real-time for purposes of clinical management and research. By comparing the analytical results that are gathered in the platform with current treatment practices, new patterns in real-time physiological data can be discovered, thus enabling earlier detection and possible prevention of various health conditions before clinical symptoms are visible. Artemis captures ECG data and ECG derived signals including the heart rate, respiration rate and chest impedance for purposes of breath detection. Other signals captured include blood oxygen saturation in addition to diastolic, systolic and mean blood pressure when such data is available [7].

Secondary usage of health data is defined as the use of personal health information collected for purposes unrelated to the initial purpose of providing direct delivery of health care to the patient/data subject. This includes activities such as research, analysis, quality and safety measurement, payment, provider accreditation and commercial activities [10]. The

secondary usage of personal health data plays an essential role in expanding current knowledge and understanding of health care and its delivery. Utilitarian motivations are strong in this research area, however, less utilitarian and more commercial focused and personally confronting issues are also significant. Privacy issues in this area are well documented and challenging [11-13]. Physiological data (e.g. ECG data) has the potential to reveal more information about an individual than what may be realized on first consideration. Nonetheless, it appears that many people are still willing to contribute the use of physiological data generated by themselves or even by their neonate(s) as a resource for the advancement of health research. Suitable privacy and governance frameworks are required in this fast moving domain. The potential for secondary use of physiological streaming data is clear and it is important for us to begin to understand what the public opinion is regarding the secondary use of their health data, specifically physiological data for health research purposes.

The remainder of this paper is structured as follows: The survey of Australian and Canadian citizens will be introduced and results pertaining to secondary use of data captured through physiological devices will be reported. This section is followed by a discussion about related work regarding neonatal contexts and associated issues surrounding the use of Big Data in health care. Prior research regarding parental attitudes towards research in the neonatal population and privacy concerns will be considered. Conclusions and future research are then presented.

Materials and Methods

The opinion of patients regarding the possibility of analytics on physiological data was explored, as part of a larger study, through a pilot survey deployed in Australia and Canada in 2009. The pilot survey included thirty attitudinal statements using a seven point Likert scale for responses and two open-ended questions. Focus groups reviewed the survey design before deployment. These groups included teenagers, aged pensioners, early school leavers, postgraduates and citizens with English as a second language. Between August and November, 482 self-administered surveys were distributed to residential blocks in regional New South Wales (NSW) and Darwin. During October and November, 250 surveys were distributed to sample populations in Ontario. The survey sampling strategy included high, medium and low socio-economic populations in regional and urban areas. Cronbach's Alpha results for reliability were in the acceptable range. Ethics approval was provided by the Research Ethics Board at the University of Ontario Institute of Technology and the Human Research Ethics Committee of the University of Wollongong.

The broader study that these questions are drawn from explored general public expectations and concerns regarding secondary use of their medical data, particularly pertaining to privacy matters [14, 15]. Constructs in the survey investigated the concepts underpinning the contemporary privacy theories of Restricted Access and Limited Control (RALC) [16] and a proposed framework for contextual integrity [17].

Results

The Australian and Canadian pilot surveys achieved response rates of 34.8% and 21.5% respectively. Question 26 in the survey explored the reuse of data captured through physiological devices: *If I was in hospital and a medical device was used to care for me – like a heart monitor or oxygen saturation monitor – I would agree for the information displayed on the screen to be saved in an anonymous way and used for medical research purposes.*

The stimulus statement used to capture feedback on physiological devices makes it clear that the data will be anonymized and used only for medical research purposes. Results for both Australians and Canadians indicate agreement with this type of secondary use of streaming data.

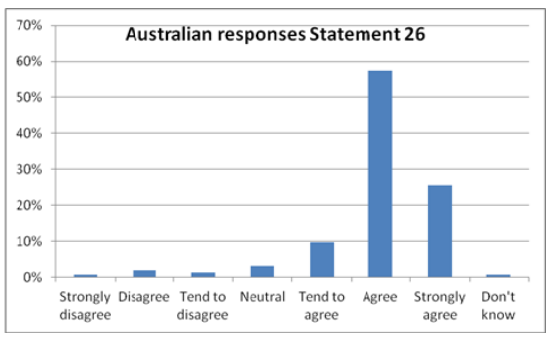


Figure 1- Frequency responses to Statement 26 on Australian surveys

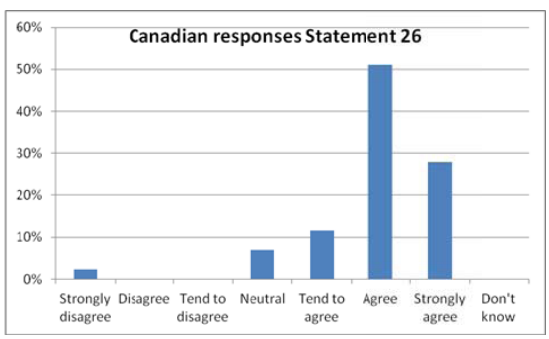


Figure 2- Frequency responses to Statement 26 on Canadian surveys

Discussion

The discussion here considers the secondary use of physiological streaming data within the neonatal intensive care context. It has been demonstrated that many parents are very willing to enroll their neonate in research studies [18, 19] even if it is known that there are significant gaps in knowledge about the study [18]. Many parents would also be willing to enroll their neonate again, if presented with the opportunity [18]. This is further supported by a study conducted by Morley et al. [20]. Parental opinions regarding the enrollment of their premature neonate(s) into several research studies in the days following birth were examined via the use of a questionnaire. Parents of preterm infants in the NICU who were invited to participate in two or more research studies were approached with this survey. Amongst the invited participants, 10% declined to allow their infants to join any studies. The majority of parents were willing to have their infant(s) be enrolled in multiple studies. In fact, 58% of the parents were willing to give their permission to enroll their infant(s) in three or more studies and 20% were willing to have their infant(s) participate in more than ten studies [20].

Many parents choose to consent for their neonate's participation in research studies because they are hopeful that it would somehow benefit their infant [18]. Many parents also want to contribute to the advancement of health research. According to findings by Morley et al. [20], 94% of the parents thought that if their baby joined a research study, the care of infants in the future would either be "better" or "very much better". Parental altruism was further demonstrated when parents were asked "Who will benefit from these studies?" in which 91% responded that "future babies", 67% said "researchers", 25% mentioned "my baby" and 2% said "no one". The insight provided by the Australian and Canadian survey results suggest that perhaps parents making decisions for neonates would also agree with the anonymised use of physiological streaming data for research purposes. This is an open research area where results could inform the governance and strategy surrounding deployment of data analytics platforms, utilizing physiological streaming data in neonatal contexts.

The issue of well informed consent arises, particularly in emotion charged contexts such as neonatal environments and the next section considers these issues.

Consent in the NICU context

The practice of acquiring informed consent is a crucial component of the research process for the protection of a neonatal research subject [21]. For consent to be considered valid, the following elements must be satisfied: full comprehension, information, and voluntariness. The participant must be mentally competent to make a free and adequately informed decision and must give their consent voluntarily and freely. Sufficient information, including the risks about the decision to be made must also be provided to the participant [22, 23].

The informed consent process is straightforward when it comes to dealing with a competent adult [24, 25]. Obtaining informed consent presents ethical and legal difficulties in certain groups of people who are considered to be part of vulnerable populations, often as a result of limited capacity or inadequate access to social goods such as rights, opportunities and power [26]. This includes but is not limited to minors and individuals living with mental disabilities or diminishing capacity [18, 22, 25, 26] as well as individuals in certain situations when one may be unable to consent for oneself. Proxy consent is therefore required for such groups of people and/or in such situations. Proxy consent is the process which occurs when individuals with the legal right to consent give advance permission to an authorized third party who is legally and competent to consent on their behalf when the individual is unable to consent for themselves. This adult may be designated through the power of attorney to consent or via a living will [22, 26, 27]. In the case of a newborn, the proxy consent at best represents parental discretion, preferences and family values [22, 25] as the neonate is incapable of communicating his/her own opinion about research and their willingness to participate [22].

It is well understood that the information provided regarding the details of a study should be sufficient enough for the reasonable parent to make an informed decision. Yet striking an ideal balance is easier said than done, since by providing too little information can render consent invalid, whereas providing too much information may consequently cause unnecessary distress. In addition to the emotional stress associated with the birth of a premature and/or critically ill infant, the mother may also have to deal with the physical stress related to the recovery period following the birth [22, 28]. The parents of such infants then face a multitude of complicated and urgent decisions while rapidly digesting new and changing information [29] in an unfamiliar environment. They are then obliged to take on the responsibility of being surrogate decision-makers on behalf of their infant. With no previous experience on which to turn, this can be a frightening experience as these parents are concerned with trying to make the right decision to benefit their infant or if they are unable to, at the very least, they want to make a decision that will benefit future infants [30].

However, it is difficult to test if parents truly understand what they are consenting their infant to when it comes to research studies [25]. Stenson et al. [31] conducted a survey that examined if parents of infants who entered into a randomized controlled trial of pulmonary function testing had any recollections about being asked to give consent for enrolling their infant in a research study and how they felt the research had affected their experience as parents of a sick infant. Although the parents were given a detailed verbal description and printed information sheet regarding the trial, of the 99 respondents, 12% could not remember being approached for consent and did not think that their infant had participated in a research study and 6% remembered being approached for consent but were unsure of whether or not their infant actually participated in a study. 89% parents who remembered being approached for consent felt that a full explanation of the studies they were enrolling their infant in had been provided to them; however only 27% and 42% of those parents felt that they understood the explanation completely and reasonably well respectively. The rest of the parents either understood a little of the explanation or not at all.

Ballard et al. [18] also examined the validity of informed consent obtained in the perinatal period in relation to their NEOPAIN study. To determine the level of parental understanding of the study, participating parents were asked open-ended questions that addressed the timing of consent, understanding of the study's purpose, benefits and risks, the voluntary nature of the project, and their willingness to enroll in future studies if applicable. Of the 64 parents who were interviewed, 5 parents (7.8%) did not remember the study or signing of consent. Of the remaining 59 parents who remembered the study, only 67.8% understood the study's purpose. It was observed that maternal understanding regarding the purpose of the study was greater than that of paternal understanding (73.3% vs 57.1%) which was particularly interesting as this study also evaluated the mother's medication effect on their memory. The medication given most frequently to the mothers in this study population was magnesium sulfate, a drug that can cause adverse effects on memory and mentation although in reality, the risks are minimal. At the time they signed their consent to enroll their infant in the NEOPAIN study, 37 of the 43 mothers were being treated with magnesium sulfate but it was noted that the administration of the drug in this case appeared to have minimal effect on the mother's ability to recall the study. It was proposed that despite exposure to labor and medication, mothers are better able to handle stress or process information more effectively. Involving the father in the consent process did not improve the overall understanding of the study or its benefits and risks.

Yet even with a double consent process, in which parents experience the consent process twice with the first time taking place before the neonate's birth and the second time occurring before the neonate's enrollment in a study, it has been found that these parents are no more likely to have given valid consent in comparison to those parents who consented only once. The reasons for this phenomenon is not well understood

although it has been suggested parental stress immediately following birth may play a role [18].

Privacy in the NICU context

Philosopher Herman Tavani provides an insightful phrase which is a useful starting point for considering privacy matters: "Privacy is a concept that is neither clearly understood nor easily defined" [32, p.11]. Within the scholarly literature, many have attempted and continue to attempt to provide the ideal definition of privacy. Some, such as Alpert (2000) see privacy as having the freedom to be whom and what one is as an individual while others such as Stephen (1873), Warren and Brandeis (1890), Westin (1967) and Gavison (1980) define it as "anything that offends decency", "being let alone", to "control over information" and "restricted access to persons and personal information" respectively, as cited by Allen [33]. There are also cultural dependencies with some cultures valuing privacy more than others [34].

Traditionally the privacy of medical patients' personal information has been protected through application of the 'limited access' theory of privacy. With the change of medium used for capture and storage of personal medical information from paper to electronic, the 'limited access' approach to privacy is under pressure due to the ease with which electronic information can be exchanged. This is an issue of growing importance with the emergence of Big Data, and the physiological streaming data available in the NICU would benefit from consideration against more contemporary privacy theories [16, 17] that go beyond the 'restricted access' or 'limited control' paradigms.

The NICU context, with volumes of streaming physiological data, is well described by Nissenbaum's definition of context:

Contexts are structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules) and internal values (goals, ends, purposes) [17, p. 132].

The survey question regarding reuse of streaming physiological data provides context related insights. The goal-ends-purposes of the data reuse were clearly described as being 'for research purposes'. The role of the survey respondent as a patient was clarified. The power structures were considered with the patient given some power to make decisions regarding the re-use of their data. The wording of the survey question implied that the clinicians were seeking shared power over the streaming data.

The deployment of privacy frameworks within the NICU to explore: (1) the enhancement of consent and simultaneously (2) privacy as contextual integrity concepts is an open research area. The early survey results and NICU specific matters considered here are a useful launch-point for further work. The broader patient privacy study referenced here explored concepts of shared power involving clinicians and patients and results indicated there was an appetite for this type of arrangement from both Australians and Canadians surveyed. There is clearly a need for the development of an appropriate patient privacy/clinician engagement model.

Biometrics from the NICU

Biometric data is considered personal information when derived from an individual to determine or verify one's identity [35]. The term "biometrics" may refer to quantifiable characteristics or the automated methods that utilize the aforementioned characteristics to identify or confirm one's identity [36]. Any human behavioural and/or physiological characteristic has the potential to be utilized as a biometric identifier provided it satisfies the criteria of universality, distinctiveness, acceptability, collectability, performance, permanence and circumvention [37]. This may have implications for the secondary use of physiological streaming data – even when the data has been anonymized.

It is unclear how biometrics captured while an individual is a patient in a NICU environment could be exploited later in that individual's life. However it is noted here that the issues surrounding biometrics will influence the future directions of secondary use of streaming physiological data.

Conclusion

This paper highlights the important contributions that physiological data, as captured by Big Data platforms, brings to health research. To date there has been little research relating to patient engagement in matters related to secondary use of such data. Contemporary privacy theories may aid navigating the emerging privacy and ethical issues, including biometrics, regarding streamlining physiological data. The survey results presented here formed part of a broader study into Australian and Canadian citizens opinions regarding application of contemporary privacy theory in medical domains. The focus here has been on the NICU context and potential for collaboration with parents of neonates on matters pertaining to consent, privacy and streamlining physiological data. The Artemis platform has been considered as one Big Data platform providing technological support.

It is important to understand the perceptions of secondary use of data in this area as early as possible and build an appropriate use model. The initial patient perceptions presented here can inform the challenging privacy aspects of a future physiological data use model. Physiological data analysis could potentially be the path to the next major advances in healthcare thus serving as a motivation to our research on a parent engaged privacy model using Big Data.

Acknowledgments

This research has been funded by the Canada Research Chairs program and received support from the Centre for Canadian-Australian Studies, University of Wollongong

References

[1] Sahoo SS, Jayapandian C, Garg G, Kaffashi F, Chung S, Bozorgi A. Heart beats in the cloud: Distributed analysis of electrophysiological "Big Data" using cloud computing for epilepsy clinical research. J Am Med Inform Assoc. 2014;21(2):263-71. doi: 10.1136/amiajnl-2013-00215.

[2] McGregor C, Catley C, James A, editors. Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit. 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2012); 2012; Rome, Italy: IEEE.

[3] McIntosh N. Clinical diagnosis of pneumothorax is late: Use of trend data and decision support might allow preclinical detection. Pediatr Res. 2000;48(3):408-15. doi: 10.1203/00006450-200009000-00025.

[4] Fabres J, Carlo W, Phillips V. Both extremes of arterial carbon dioxide pressure and the magnitude of fluctuations in arterial carbon dioxide pressure are associated with severe intraventricular hemorrhage in preterm infants. Pediatrics. 2007;119(2):299-305. doi: 10.1542/peds.2006-2434.

[5] Tuzcu V, Nas S, Ulusar U. Altered heart rhythm dynamics in very low birth weight infants with impending

intraventricular hemorrhage. Pediatrics. 2009;123(3):810-5. doi: 10.1542/peds.2008-0253.

[6] Shankaran S, Langer J. Cumulative index of exposure to hypocarbia and hyperoxia as risk factors for periventricular leukomalacia in low birth weight infants. Pediatrics. 2006;118(4):1654-9. doi: 10.1542/ peds.2005-2463.

[7] McGregor C. Big data in neonatal intensive care. Computer. 2013;46(6):54-9. doi: 10.1109/MC.2013.157.

[8] Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs. 2014;33(7):1123-31. doi: 10.1377/hlthaff.2014.0041.

[9] boyd D, Crawford K. Six provocations for big data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society; 21 September 2011; Oxford. 2011.

[10] Safran C, Bloomrosen M, Hammond W, Labkoff S, Markel-Fox S, Tang P, et al. Toward a national framework for secondary use of health data: An American medical informatics society white paper. 2006.

[11] Lowrance W. Learning from experience: Privacy and the secondary use of data for health research. London: The Nuffield Trust, 2002.

[12] Willison D. Privacy and the secondary use of data for health research: Experience in Canada and suggested directions forward. J Health Serv Res Policy. 2003;8(supplement 1):17-23. doi: 10.1258/135581903766468837.

[13] Willison D, Steeves V, Charles C, Schwartz L, Ranford J, Agarwal G, et al. Consent for use of personal information for health research: Do people with potentially stigmatising health conditions and the general public differ in their opinions? BMC Med Ethics. 2009;10(10). doi: 10.1186/147269391010.

[14] Heath J. Emerging consumers view of secondary uses of medical data. In: Michael K, editor. International Symposium on Technology and Society; Piscataway, New Jersey, USA: IEEE; 2010. p. 87-95.

[15] Heath J, Dawson L, Pearson R, editors. Consumer perceptions of medical data ownership and sharing for secondary use purposes. HIC Big Data 2013 Conference; 2013 18-19 April; Melbourne, Australia.

[16] Tavani H, Moor J. Privacy protection, control of information and privacy-enghancing technologies. SIGCAS Comput Soc. 2001;31(1):6-11. doi: 10.1145/572277.572278.

[17] Nissenbaum H. Privacy in context: Technology, policy and the integrity of social life. Stanford, California: Stanford University Press; 2010.

[18] Ballard H, Shook L, Desai N, Anand K. Neonatal research and the validity of informed consent obtained in the perinatal period. J Perinatol. 2004;24:409-15. doi: 10.1038/sj.jp.7211142.

[19] Stenson B, Becher J-C, McIntosh N. Neonatal research: the parental perspective. Arch Dis Child Fetal Neonatal Ed. 2004;89(4):321-4.

[20] Morley C, Lau R, Davis P, Morse C. What do parents think about enrolling their premature babies in several research studies? Arch Dis Child Fetal Neonatal Ed. 2005;90(3):F225–8.

[21]Burgess E, Singhal N, Amin H, McMillan DD, Devrome H. Consent for clinical research in the neonatal intensive care unit: A retrospective survey and prospective study. Arch Dis Childhood Fetal Neonatal Ed. 2003;88(4):280-6.

[22] Mason S. Obtaining informed consent for neonatal randomised controlled trials - an "elaborate ritual"? Arch Dis Child Fetal Neonatal Ed. 1997;76(3):F143–5. doi: 10.1136/fn.76.3.F143.

[23] Oberle K, Singhal N, Huber J, Burgess E. Development of an instrument to investigate parents' perceptions of research with newborn babies. Nurs Ethics. 2000;7(4):327–38.

[24] Committee on Bioethics American Academy of Pediatrics. Informed consent, parental permission, and assent in pediatric practice. Pediatrics. 1995;95(2):314–7.

[25] Cooke R. Good practice in consent. Seminars in Fetal & Neonatal Medicine. 2005;10(1):63–71.

[26] Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-council policy statement: Ethical conduct for research involving humans. 2 ed. Ottawa, Canada. 2014.

[27] Richards E. Proxy consent 1993 [cited 2014 May 16]. Available from: http://biotech.law.lsu.edu/Books/lbb/x302.htm.

[28] Singhal N, Oberle K, Burgess E, Huber-Okrainec J. Parents' perceptions of research with newborns. J Perinatol. 2002;22(1):57–63. doi: 10.1203/00006450-199904020-00214.

[29]Baker B, McGrath JM. Engaging Families in Neonatal Research. Newborn Infant Nurs Rev. 2010;10(1):5-7. doi: 10.1053/j.nainr.2009.12.004.

[30] Golec L, Gibbins S, Dunn MS, Hebert P. Informed consent in the NICU setting: an ethically optimal model for research solicitation. J Perinatol. 2004;24(12):783–91. doi: 10.1038/sj.jp.7211198.

[31] Stenson B, Becher J, McIntosh N. Neonatal research: The parental perspective. Arch Dis Child Fetal Neonatal Ed. 2004;89(4):F321-4. doi: 10.1136/adc.2002.021931.

[32] Tavani H. Privacy online. SIGCAS Comput Soc. 1999;29(4):11-9. doi: 10.1145/572199.572203.

[33] Allen A. Genetic privacy: Emerging concepts and values. In: Rothstein MA, editor. Genetic secrets: Protecting privacy and confidentiality in the genetic era. New Haven, CT: Yale University Press; 1997. p. 31–59.

[34] Moor J, editor. Towards a Theory of Privacy for the Information Age. Amherst, New York: Prometheus Books; 2000.

[35] International Biometric Group LLC. BioPrivacy FAQ's and definitions [updated 2010; cited 2014 Apr 8]. Available from: http://www.bioprivacy.org/.

[36] National Science Technology Council [NSTC] Committee on Technology, Committee on Homeland and National Security SoB. Privacy & biometrics: Building a conceptual foundation. 2006. p. 1–57.

[37] Kannavara R, KL S. Topics in biometric human-machine interaction security. IEEE Potentials. 2013;32(6):18–25. doi: 10.1109/MPOT.2013.2248891.

Address for correspondence

University of Ontario Institute of Technology
2000 Simcoe Street North, Oshawa, Ontario, Canada
Email: c.mcgregor@ieee.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-458

Driving the Profession of Health Informatics: The Australasian College of Health Informatics

Christopher Pearce[a], Klaus Veil[a], Peter Williams[a], Andrew Cording[a], Siaw-Teng Liaw[a], Heather Grain[a]

[a] *Council of the Australasian College of Health Informatics*

Abstract

Across the world, bodies representing health informatics or promoting health informatics are either societies of common interest or universities with health informatics courses/departments. Professional colleges in Health Informatics (similar to the idea of professional colleges in other health fields) are few and far between. The Australasian College of Health Informatics has been in existence since 2001, and has an increasing membership of nearly 100 fellows and members, acting as a national focal point for the promotion of Health Informatics in Australasia.

Describing the activities of the college, this article demonstrates a need for increasing professionalization of Health informatics beyond the current structures.

Keywords:

Professional education; Professional Review Organisations.

Introduction

Health Informatics (HI) as a distinct entity is an evolving field [1], and has developed from a craft based solution to the problems of computerising the health sector, into a distinct field of practice. The next step is to understand HI as a profession, one in which a 'Health Informatician' can be seen as part of a specialised group, in the same way that, for example, surgeons identify as a professional group.

The beginnings of Health Informatics began in the 1960's with the development of the concept of information systems in medicine, both in the US and the UK. The International Medical Informatics Association was formed in 1989, and in 1994 the first MEDINFO was held [2]. It is therefore over this time that the discipline of informatics evolved, as participants grappled with the implementation of a game-changing technology in medicine [3].

In order to be considered a separate academic discipline, a field of study needs several characteristics [4], sufficient to be:

1. A particular object of research (e.g. law, society, politics), though the object of research maybe shared with another discipline;
2. A body of accumulated specialist knowledge referring to their object of research, which is specific to them and not generally shared with another discipline;
3. Have theories and concepts that can organise the accumulated specialist knowledge effectively;
4. Use specific terminologies or a specific technical language adjusted to their research object;
5. Have developed specific research methods according to their specific research requirements, and maybe most crucially; and
6. Must have some institutional manifestation in the form of subjects taught at universities or colleges, respective academic departments, and professional associations connected to it.

With the development of a significant research base, journals devoted to the discipline, Health Informatics qualifies on all these counts. Therefore, the academic basis for HI is relatively strong, particularly in the US.

Does the same apply for HI as a profession? The features of a profession have some similarities and some differences to academic standing. A profession must have [5]:

1. A systematic body of knowledge and theory: in other words have those features of an academic discipline outlined earlier;
2. Professional authority: which in turn derives from the first point. To have specialised knowledge and understanding allows a profession to have deemed or delegated authority over that body of theory;
3. The sanction of the community: a profession seeks to have control over aspects of its internal functioning – training, standards, and regulation – all of which must come from persuading the community that the profession can and should have control indeed, that it is in the communities (or the public) interest in having such control;
4. A regulatory code of ethics: where a profession exists, it effectively functions as a monopoly, or a dominating interest in the field. The profession activities must therefore be regulated, and self-regulation is first by a code of ethics, and second, by structures that mandate standards of entry; and
5. A professional culture: the establishment and maintenance of a professional community in which to set and monitor standards, maintain skill sets, and drive the growth of the profession.

It is in these latter areas that the ACHI as a professional college sits, and where ACHI derives its professional authority from – to drive the profession of HI in Australia.

Across the world, bodies representing health informatics or promoting health informatics are either societies of common interest, universities with health informatics courses/departments, or academic professional bodies. These societies of common interest are represented by organisations such as the American Medical Informatics Association (AMIA), the Health Informatics Society of Australia (HISA), or the Colombian Association of Health Informatics, for example. These organisations generally have no barriers of entry, and have as their purpose the promotion of activities within their members. The International Medical Informatics Association (IMIA) lists 57 such organisations in its membership.

Universities with health informatics faculties are characterised by their focus on research and teaching in that domain. Teaching staff are appointed for their demonstrated domain knowledge, researchers for their demonstrated publication outputs, and students for their demonstrated prospects in achieving health informatics competencies.

A third model exists, which is an academic professional body dedicated to standards and education. This model is common in medicine, and acts as a supplement to the other two. The Royal College of General Practice (RCGP) is an example: 'We are the professional membership body and guardian of standards for family doctors in the UK, working to promote excellence in primary healthcare'. The RCGP sits alongside the British Medical Association and works in conjunction with the various academic departments to further the discipline. Professions are ideally placed to work with government on matters of standards and training [6].

A defining characteristic of these bodies is a standard of entry, acquired by 'fellows' after an approved period of training and demonstrating the achievement of a certain standard. These bodies are then active in the defining of standards of practice, revalidation, and policy surrounding the scope of activities. Their main role is therefore to bridge the gap between academic practice and professional practice.

Such a model is rare in health informatics. The American College of Medical Informatics is an example for health biomedical informatics (http://www.amia.org/programs/acmi-fellowship). The ACMI model was used to guide the establishment of the ACHI.

This paper describes the Australasian College of Health Informatics (ACHI), a unique body in this area. ACHI is:

'*the professional body for Health Informatics in the Asia-Pacific Region. The credentialed Fellows and Members of the College are national and international experts, thought leaders and trusted advisers in Health Informatics. ACHI sets standards for education and professional practice in Health Informatics, supports initiatives, facilitates collaboration and mentors the community'*.

It is the only body of its type that is a member of the International Medical Informatics Association (IMIA). The ACMI is a member through its being an integral part of the AMIA.

The college was established according to a perceived need, to have a qualification of sufficient standard to demonstrate that the holder is an established professional, worthy of developing and implementing the evidence base in HI. Again, using the model of the medical colleges, a surgeon must have achieved the fellowship of the college of surgeons before being able to practice in that field.

In Australia, formal training in Health Informatics (HI) is rare, with few designated courses and variable support from universities. Many active in large projects have been either 'enthusiastic amateurs' who trained in another profession, or have learned on the job through a variety of methods. Fellows have come through the college with demonstrable Health Informatics PhD's, awarded through engineering departments, general practice, anaesthesia, and even business faculties. Others have come through paths involving mostly industry based training, with an emphasis on practical implementations.

A not unusual scenario is of a clinician thrust into an informatics role to help implement a new clinical system, and having to understand clinical coding, adoption and workflow theories, all in a rapid timeframe. Learning paths are therefore driven by interest and need, rather than following a formal path.

ACHI was founded in 2001. Professor Enrico Coiera was the first president. Enrico was an early champion of 'medical informatics' [7] in Australia. ACHI currently has 85 fellows and 11 members. The majority are from Australia, but ~10% are from New Zealand and another 10% are based overseas (Hong Kong, Japan, Malaysia, England, and Germany). The overall membership has doubled in the past 18 months, suggesting a significant demand for a professional body of this type, and 20% are female.

While many are in university departments with clear academic roles, the majority are in industry, working for government departments, hospitals, Medicare Locals (primary care support organisations), and companies. Many work in isolation, in that they are the only college fellow/member in that organisation or department.

Membership

There are two levels within the college: fellows and members. In the absence of a formal structured education program in Health Informatics, applicants have to demonstrate skills and expertise in the field, by applying to the membership committee. To join ACHI, an applicant must:

- Have made a (distinguished for fellows, measured for members) contribution to the field of Health Informatics;
- Have a minimum of eight years (for fellows) or five years (for members) of practical experience in Health Informatics (excluding time spent as an undergraduate student). For fellows, at least five must have been in a position from which the actions of the nominee have resulted in outcomes which can be seen to support the 'distinguished contribution' to the field in Australasia; and
- Be an ethical user of health information, demonstrate professional integrity, and have their academic documentation verified and accepted.

To achieve the required level, members must accrue 60 points, and fellows 100 points of recognised activities. As examples, a PhD in a relevant discipline is worth 50 points, a research paper 20 points, each year in a HI post 5 points, and chief investigator in a significant research project 20 points.

Governance

ACHI is governed by a seven member council, with four committees reporting to council. They are:

- Membership committee: receives and determines membership applications, and developing a reaccreditation program;
- Education committee: developing models of career pathways and course accreditation programs;
- Program evaluation committee: tasked with performing rapid evidence reviews to inform policy development. The last two areas have been on personal health records and consumer access to diagnostic testing results; and
- Professional standing committee: tasked with developing a plan to promote heath informatics as a discipline to potential stakeholders (government, health institutions).

Figure 1 – Governance Structure

Activities

The activities of the college for the next three years are guided by its strategic plan, focussed on several areas.

Figure 2 – Priority Areas

Education

Here, the objective is to see that the College sets the framework for education excellence in Heath Informatics in the Asia Pacific region. In addition to the education committee's role in developing a vision for a career path in Health Informatics, taking into account the diversity of entry points and education provision, including the Certified Health Informatician of Australia program (CHIA - an entry level practical qualification), the college will strengthen links with international organisations (including AMIA, IMIA, and EFMI) active in the area, particularly in the field of course accreditation.

At the moment, there is no national standard to inform course development, and therefore no driver for employing organisations to request appropriate skills.

Membership

Designed to foster a community of health informatics professionals, this inwards looking target is designed to increase membership, establish connections with international bodies, and develop a framework for revalidation. Revalidation is an essential feature of a professional college if it is to be able to demonstrate that its fellows remain skilled. In health, revalidation is now compulsory, and health informatics should be no exception.

An important advancement is the establishment of a mentoring program. Prospective fellows who do not meet the criteria can be allocated to a fellow to advise them on career development.

Advocacy

Crucial to the activities of a professional college is to be able to advocate on behalf of its members, and ACHI regularly contributes to the policy discourse on Health Informatics and is seen as an expert in the area. These papers are all available through ACHI's website <www.ACHI.org.au>. Despite having fellows who are national and international experts in various fields, their voice has often been silent in policy development in Australia, particularly at a time when, nationally, the government is embarking on an ambitious shared EHR project [8].

The increased activity of ACHI is correcting this balance. Governments develop stakeholder consultation plans, and the existence of a formal college has allowed fellows to be represented at the table alongside with other professional colleges.

The activities of the Program Evaluation Committee have been crucial in this area, allowing the college to present evidence based views to inform policy, rather than vested or stakeholder views.

Engagement

Alongside advocacy, engagement broadly across the professional spectrum is an activity designed to raise the profile of informatics as a profession. This means not only engaging with HISA and similar bodies, but also with the other health professional colleges – surgeons, physicians, and the like. Clinical engagement is crucial to the success of health projects, but that clinical involvement must be tempered by the appropriate level of Health Informatics professionalism [9].

Professional Recognition

The final piece of the puzzle, and underpinning all other activities, the role of ACHI is to see that the profession of health informatics be duly recognised by all, and that the College be seen as the standards and expert body in the field.

This area has many aspects, ranging from the bureaucratic to the broad. Health informatics is not represented in the statistical classification of professions, for instance [10]. Nor is HI and the presence of the college well understood by other bodies. Reconciling the definitional issues between the pragmatic, industry based definition of Health Informatics versus the academic definitions (biomedical informatics/medical informatics) [11] is another issue.

Discussion

For HI to be considered a true profession, there needs to be bodies worldwide such as ACHI to promote the profession. Associations, with no barriers of entry and no standards have a role in developing a community, but only a professional body can have the moral authority and standing to negotiate and maintain standards [12]. In this light, ACHI's role in supporting the profession becomes clear, revisiting Greenwood's criteria:

1. A systematic body of theory: well defined through existing academic work, but supported through the community created by ACHI, with meetings, the 'electronic Journal of Health Informatics' (www.eJHI.net), and other activities;
2. Professional authority: this is a key goal for ACHI. In an industry environment where HI is not necessarily recognised as such, the advocacy and engagement areas of the strategic plan are designed to develop and improve the authority held by the profession;

3. The sanction of the community: again, a key goal. In order to be able to influence the professional development, ACHI must persuade the 'community' that this is in the community's interest. Setting the educational and training standards is the essential function, as is the engagement and advocacy roles;
4. Regulatory code of ethics: ACHI has such a code. Designed to set out the standards of behaviour that are required of health informatics professionals in Australia. It deals with all aspects of professional activity including ACHI members' duties to patients, the public, employers and colleagues. Any ACHI member who fails to comply with these standards may be removed as an ACHI member. The code of conduct can be found at: <http://www.ACHI.org.au>; and
5. Professional culture: this is in many ways the most important, and the most complex one in the early stages of development of the profession. Fellows came from a wide variety of backgrounds and experiences. They can be medical professionals well steeped in medical ethics, or from other backgrounds with different ethical frames [13]. It is the most important aspect of creating the community of the profession. Approximately 15% of the college are active on committees and in developing positions, and the College has an active email group.

The progress of the college over the last twelve to eighteen months has been encouraging. Membership has doubled, engagement has increased both within and without. ACHI is poised to develop from a small group to a fully-fledged professional body influencing the development of the profession.

Challenges

ACHI is beginning to form an important role in developing the profession in the region. Although its main impact is in Australia, its role is developing outside the country, in particular in New Zealand. The challenges it faces are both internal and external.

Internal

The demands on the college far exceed its capacity to achieve them. Especially at this stage of development, the college employs no staff, and all work done is in a voluntary capacity by people who are already, by the very nature of the field, very busy. This limits the ability to deliver many of the required activities.

The heterogeneity of fellow development paths also means that there are a high variety of views existing within the profession. These vary from basic definitional issues mentioned earlier to discussions on appropriate role for the college.

External

The most significant issue for the college and its external relationships devolves primarily from the first challenge. As the presence of the college becomes more prominent, so does the external expectations of capacity and ability to deliver. In an all-volunteer body with a relatively small number of fellows many of them new to the profession, there is significant risk of failure to deliver on expectations, and for this reason, expectation management is an important area.

Conclusion

This paper has outlined the activities of ACHI, a professional body for health informatics. More importantly, it has outlined the framework for a profession, demonstrated Heath Informatics can be seen to be a profession. The growth of ACHI demonstrates an unmet need, not just in our local area, but one that needs to be replicated elsewhere for HI to achieve full recognition, and its practitioners to receive the recognition they deserve.

Acknowledgements

We acknowledge Associate Professor Chris Bain, for his thoughtful comments, and all fellows and members who have contributed to the college's development. We also acknowledge all previous presidents.

References

[1] Oh H, Rizo C, Enkin M, and Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005: 7(1):e1.

[2] Conrick M. Health Informatics, Transforming Healthcare with Technology. South Melbourne: Thomson Social Science Press, 2006.

[3] Blumenthal D, and Glaser JP. Information technology comes to medicine. N Engl J Med 2007: 356(24): 2527-34.

[4] Goodlad S. What is an Academic Discipline? . In: Cox R, ed. Cooperation and Choice in Higher Education. 1. London: University of Londin Teaching Methods Unit, 1979; pp. 11.

[5] Greenwood E. Attributes of a Profession. Social Work 1957: 2(3): 194-200.

[6] Dwan K, and Boyce R. Competition Policy and Intra-professional conflict: re-regulating general practice. International journal of Public Sector Management 2003: 16(2): 141-52.

[7] Coiera E. Medical informatics. BMJ 1995: 310(6991): 1381-7.

[8] Pearce C, and Bainbridge M. A personally controlled electronic health record for Australia. J Am Med Inform Assoc 2014: 21(4): 707-13.

[9] Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, and et al. The impact of clinical leadership on health information technology adoption: systematic review. Int J Med Inform 2014: 83(6): 393-405.

[10] Martin-Sanchez F, and Gray K. Recognition of health informatics in Australian standard classifications for research, occupation and education. Stud Health Technol Inform 2014: 204: 92-7.

[11] Venot A, Burgun A, Despres S, and Degoulet P. Medical Informatics as a scientific discipline. In: Venot A, Burgun A, Quantin C, eds. Medical Informatics, e-Health Fundamentals and Applications. Paris: Sprigner-Verlag, 2014; pp. 12-21.

[12] Starr P. The social transformation of American medicine. New York: Basic Books, 1982; xiv..

[13] McCrickerd J. Metaphors, models and organisational ethics in health care. J Med Ethics, 2000: 26(5): 340-5.

Address for correspondence

Associate Professor Chris Pearce. Inner East Melbourne Medicare Local, Melbourne, Australia cpearce@iemml.org.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-462

Person-Specific Standardized Vulnerability Assessment in Health and Social Care

Léa A. Deleris[a], Pól Mac Aonghusa[a], Robert Shorten[a]

[a] IBM-Research – Ireland, Dublin, Ireland

Abstract

We describe an integrated person-specific standardized vulnerability assessment model designed to facilitate patient management in health and social care. Such a system is not meant to replace existing health and social assessment models but rather to complement them by providing a holistic picture of the vulnerabilities faced by a given patient. In fact, it should be seen as a screening tool for health and social care workers. One key aspect of the modeling framework is the ability to provide personalized yet standardized multi-dimensional assessments of risk based on incomplete information about the patient status, as is the case in screening situations. Specifically, we integrate a Markov chain model describing the evolution of patients in and out of vulnerable states over time with a Bayesian network that serves to customize the dynamic model. We present an application in the context of elder care.

Keywords:

Comprehensive Health Care; Patient-Centered Care; Aging; Risk Factors; Vulnerability.

Introduction

Health and social programs represent a significant, and growing, proportion of the public sector budget across OECD economies. In fact, the OECD reported that public sector budgets devoted to social programs across OECD economies represented 22% of GDP [1]. In the US, Medicaid spending ($457 billion in 2014) is expected to almost double by 2020 [2]. As expenses continue to grow, it has become essential for sustainability to look into solutions for reducing health and social care expenses while possibly minimizing the effect on the support provided (if not to improve its quality at the same time). Such an objective is not necessarily as utopic as it sounds. Indeed, there remains ample room for improvement within the current delivery of health and social services that mostly operate as a set of side-by-side, one-size-fits-all processes (unemployment, child support, housing support…). In addition, those services have typically been designed to address simple situations and tend to be ill-suited for complex ones; i.e., when a person presents a variety of health and social needs. Unfortunately, such complex situations are often associated with high costs. In New York State (NYS), about five percent of Medicaid beneficiaries account for more than half of Medicaid spending [3] and in Camden, NJ, one percent of patients account for a third of the city's medical costs [4]. Hence, finding ways to (i) better coordinate the actions of social and health services across silos, and (ii) proactively identify situations that can become high-cost, high-need are promising avenues to address the health and social care spending challenge.

Our proposed model seeks to help in those two directions at an operational level. Specifically, the objective of our model is to assist care coordinators (whether healthcare or social care) as they receive new situations or update existing ones. The model provides them with a holistic high-level assessment of the clients' current needs, along with assessments of the vulnerabilities (upcoming problems) of those clients. With this decision support information, careworkers are able to effectively address current situations and to identify potentially complex and costly ones ahead of time.

Note that the benefits of looking at holistic assessment of needs, in particular across health and social issues, reach beyond the efficiency gains of improved coordination. In fact, the main reason for taking a holistic perspective is the existence of strong cross-influences between health and social factors, often termed as the social determinant of health [5]. For instance, research has shown that an additional four years of education, besides lowering mortality, reduces the risk of heart disease by 2.16 percent and the risk of diabetes by 1.3 percent [6]. Understanding and accounting for the cross-influences among domains is critical in the effort to better manage patients' wellness altogether and to reduce health inequalities that arise between people of different socio-economic or ethnic groups.

Methods

Our vulnerability modeling framework consists of a single integrated model that evaluates a set of vulnerability metrics across several domains; for instance, "Physical Health", "Mental Health", "Income", "Shelter", and "Fundamental Activities of Daily Living" in the context of eldercare.

Dynamic Model through Markov Chain

The heart of a vulnerability model is a dynamic model that represents the probabilistic evolution of a person over time. From that model, and given the current state of a person, we can estimate the chance of that person reaching a vulnerable state. Specifically, we use a homogeneous Markov chain to represent the transitions of persons through states, a subset of which have been identified as vulnerable states. A Markov chain is a random process that represents the sequence of states visited by a system over time. It is characterized by the Markov property (also called "memoryless") which stipulates that the evolution of the system only depends on its current state and not on how it reached that state. This property enables one to characterize the model with few parameters, and simplifies computations. In a simple two-state chain, for instance, "Employed" and "Not Employed", the only two parameters that need to be provided are (i) the probability of becoming unemployed in the next time period given one is currently employed and (ii) the probability of becoming employed given one is currently unemployed.

As our goal is to consider multiple dimensions of vulnerability at the same time, the Markov chain needs to consider state combinations, as in the example in Figure 1, which looks at Shelter and Employment vulnerabilities, albeit in a very crude way (modeling two states for each domain):

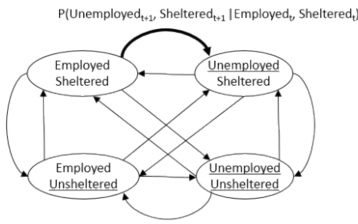


Figure 1– Simple Markov Chain Model to Represent Evolution of a Person over time

Note: To make the diagram readable, we have not represented the self-transitions. The transition probability that is spelled out is associated with the bolded arc.

Based on the Markov chain parameters, it is possible to compute future-behavior statistics aimed at characterizing vulnerability. We have chosen to rely on the mean first passage time (MFPT) to a vulnerable state conditional on the current state. In our example, assuming the person's current state is "Employed, Sheltered", then the Employment vulnerability would be the average time of the first "visit" to either the "Unemployed, Sheltered" or "Unemployed, Unsheltered" states. MFPT provides a characterization of the short-term behavior of the model and is thus particularly suited as a measure of immediate vulnerability of a person.

Mathematically, Let *d* index vulnerability domains and let D denote the number of such domains. Each domain is described by a state space $\boldsymbol{S^d}$ where $\boldsymbol{S'^d}$ is the subset of $\boldsymbol{S^d}$, which corresponds to those deemed vulnerable. In our preceding example, we would have D = 2. The first domain is "Shelter" with **state space** $\boldsymbol{S^1} = \{\boldsymbol{Sheltered, Unsheltered}\}$, **and vulnerable state set** $\boldsymbol{S'^1} = \{\boldsymbol{Unsheltered}\}$, and the second dimension is **"**Employment**"**, with **state space** $\boldsymbol{S^2} = \{\boldsymbol{Employed, Unemployed}\}$, **and vulnerable state set** $\boldsymbol{S'^2} = \{\boldsymbol{Unemployed}\}$. We denote by $\boldsymbol{S}$ the product space of each state space for each dimension. $\boldsymbol{S} = \boldsymbol{S^1} \times \cdots \times \boldsymbol{S^D}$. Let N denote the total number of states, $\boldsymbol{N} = \boldsymbol{card}(\boldsymbol{S})$.

Let $\boldsymbol{x_t^d}$ represent the state of a person at discrete time intervals *t* along vulnerability domain *d*. Thus, $\boldsymbol{x_t} = (\boldsymbol{x_t^1}, \cdots, \boldsymbol{x_t^D})$ captures the comprehensive state of the person at time *t*. We assume that there exists a homogeneous Markov chain with transition matrix $\boldsymbol{M}$ where $\boldsymbol{m_{ij}} = \boldsymbol{P}(\boldsymbol{x_{t+1}} = \boldsymbol{s_j} | \boldsymbol{x_t} = \boldsymbol{s_i})$ where $\boldsymbol{s_i}, \boldsymbol{s_j} \in \boldsymbol{S}$, which describes the evolution of a person across states. Finally, let $\boldsymbol{\mu_{ij}}$ denote the mean first passage time $\boldsymbol{\mu_{ij}}$ from state $\boldsymbol{s_i}$ to state $\boldsymbol{s_j}$, defined as the expected number of steps to arrive at destination $\boldsymbol{s_j}$ for the first time starting from state $\boldsymbol{s_i}$. For each domain *d*, we define the vulnerability index associated with current state $\boldsymbol{s_i}$ as follows:

$$\begin{cases} \boldsymbol{v_i^d} = \boldsymbol{\min}_{\boldsymbol{j}:\boldsymbol{s_j}^d \in \boldsymbol{S'^d}} \boldsymbol{\mu_{ij}} & \forall \boldsymbol{i}: \boldsymbol{s_i}^d \notin \boldsymbol{S'^d} \\ \boldsymbol{v_i^d} = \boldsymbol{0} & \forall \boldsymbol{i}: \boldsymbol{s_i}^d \in \boldsymbol{S'^d} \end{cases}$$

Indeed, if the current state does not correspond to a vulnerable state for domain *d* ($\boldsymbol{s_i}^d \notin \boldsymbol{S'^d}$), then we are interested in the mean first passage time to any destination state for which its *d*th component is deemed vulnerable, hence the need to use the minimization operator. When the starting state corresponds to a vulnerable state ($\boldsymbol{s_i}^d \in \boldsymbol{S'^d}$), the vulnerability index is is set to 0 to indicate current need. Finally, we call the vector $\boldsymbol{V_i} = (\boldsymbol{v_i^1}, \cdots, \boldsymbol{v_i^D})$ the vulnerability profile of the person corresponding to current state $\boldsymbol{s_i}$.

Bayesian network for standardised customization

Observe that if the parameters of the Markov chain are fixed, any person in a similar current state would have the exact same vulnerability profile, providing minimal differentiating information. To allow for model customization, we have chosen to let the parameters of the Markov chain depend on a set of relevant factors. Specifically, we assume that there exists an underlying Bayesian network that articulates the relationships among the factors themselves and also between the factors and the state of a person. Bayesian networks are graphical models that represent relationships among uncertain variables [7]. The nodes represent the variables and the arcs represent the conditional dependencies in the model. Markov chains and Bayesian networks are consistent modeling frameworks in the sense that they are both based on representing conditional probabilities.

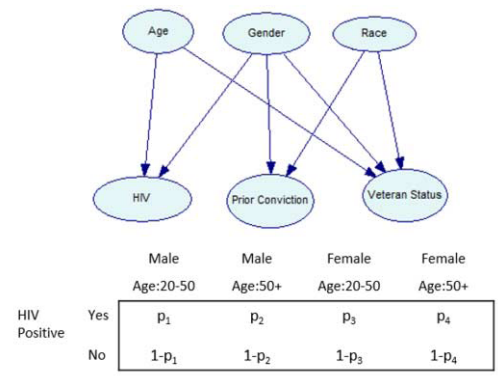


Figure 2 – Illustrative Bayesian Network

Figure 2 shows an example of such a Bayesian network with six factors. This model implies that *Age*, *Gender,* and *Race* are independent of one another but that *Veteran Status* is influenced by all three. In addition to the structure of the model, Bayesian networks are characterized by a quantitative layer that represents the distributions of the variables conditional on all possible parent scenarios. At the bottom of Figure 2, we present the conditional probability table for the variable *HIV positive,* where p_2 represents the probability of being HIV positive given one is at least 50 years old and male. Bayesian networks are especially efficient at computing inference queries; i.e., conditional probabilities involving the subsets of the variables in the network. In our example such a query could be asking for the probability of a person being *HIV positive* given *Gender* and *Veteran Status*.

Note that we can extend such a model to include variables corresponding to a Markov chain model as presented in Figure 3. By extending the Bayesian network so as to pair it with the Markov chain model, inference computations can provide personalized parameters for the Markov chain, accounting for the current knowledge of a person's situation, both current state and known factors, as large or limited as this knowledge may be.

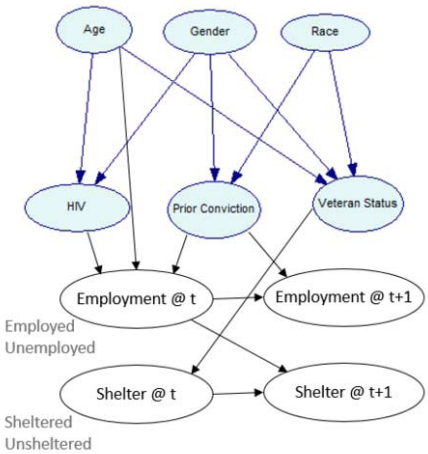


Figure 3– Illustrative Paired Bayesian Network

Mathematically, Let $\boldsymbol{\chi} = \{\boldsymbol{X_1}, \cdots, \boldsymbol{X_n}\}$ denote the set of all the variables represented in the Bayesian network. We decompose $\boldsymbol{\chi}$ into three groups:

- $\mathcal{F} = \{ x_k^F : k = 1, \cdots, K\}$: set of variables representing explanatory factors
- $\mathcal{S}_t = \{ x_t^d : d = 1, \dots, D\}$: set of variables representing the state variables at current time t
- $\mathcal{S}_{t+1} = \{ x_{t+1}^d : d = 1, \dots, D\}$: set of variables representing the state variables at future time t+1

The last two sets enable the pairing to the Markov chain model. Indeed, each element in the set $\boldsymbol{\mathcal{S}_t}$ corresponds to one of the vulnerability domains and $\boldsymbol{Val}(\boldsymbol{x_t^d}) = \boldsymbol{Val}(\boldsymbol{x_{t+1}^d}) = \boldsymbol{\mathcal{S}^d}$. Where in a Markov model a state s was represented by one (vector) variable, it is represented by a set of variables in the Bayesian network. For any observed evidence $\boldsymbol{\mathcal{E}}$ belonging to a subset of the state space of $\boldsymbol{\mathcal{F}}$, and any state $\boldsymbol{s_i}, \boldsymbol{s_j} \in \boldsymbol{\mathcal{S}}$, the Bayesian network model can provide through inference the value of the parameters of the Markov chain model :

$$\boldsymbol{P}\left(\boldsymbol{x_{t+1}^1 = s_j^1}, \dots, \boldsymbol{x_{t+1}^d = s_j^d}, \dots, \boldsymbol{x_{t+1}^D = s_j^D} \middle| \boldsymbol{x_t} \ = \boldsymbol{s_i} \ , \boldsymbol{\mathcal{E}}\right)$$

Therefore, when we let $\boldsymbol{\mathcal{E}}$ represent the observed information about a person related to factors in $\boldsymbol{\mathcal{F}}$, we can estimate the associated personalized Markov chain parameters and thus that person's vulnerability profile. Finally, as knowledge about a person's situation evolves, the vulnerability indexes evolve as well.

Characteristics of the modeling framework

This modeling framework presents several advantages. Through the Bayesian network model, the vulnerabilities across domains are made interdependent. In particular, in the context of health and social care, this means that the model can capture the *interactions* among social factors, health factors and vulnerabilities. The resulting vulnerability profile is *personalized yet standardized*. It is personalized as the Bayesian network provides a way to determine the parameters of the Markov chain model based on observed state of the person. It is standardized, as the same customization model is used for everyone. Note also that the vulnerability indexes are associated with a *physical* (thus interpretable) measure, in the form of average amount of time. This consistent use of one metric enables both within-subject and across-subject comparisons. Finally, the model is designed to function with *incomplete information* about a person's current state and profile, a key difference with assessment tools, making it a useful complement as a screening step.

In the presentation of the model above, we have focused on measuring vulnerability through the risk of occurrence of a vulnerable state. The model can be used at the same time to evaluate severity upon occurrence. Specifically, for a vulnerability domain d, for a state $\boldsymbol{s_i}^{\boldsymbol{d}}$ corresponding to a vulnerable state, the severity index $\boldsymbol{w_i}^{\boldsymbol{d}}$ is defined as the mean first passage time to a non-vulnerable state $\{\boldsymbol{s} \in \boldsymbol{\mathcal{S}}: \boldsymbol{s^d} \notin \boldsymbol{\mathcal{S}'^d}\}$. For a state $\boldsymbol{s_i}^{\boldsymbol{d}}$ that does not correspond to a vulnerable state, we can either set $\boldsymbol{w_i}^{\boldsymbol{d}} = \boldsymbol{0}$, or envision estimating severity by how long, on average, one would stay vulnerable if one were to become vulnerable.

Besides the vulnerability profile, the model yields further insights into a person's situation. In particular, we can provide (i) guidance into what additional information could be useful (informative factors) and (ii) explanations (influential factors) by identifying which aspects of the person profile strengthen and weaken his/her profile. Both are obtained by performing one-way sensitivity analyses. For instance, let's consider domain d, and current vulnerability index $\boldsymbol{v_i^d}(\boldsymbol{\mathcal{E}})$, corresponding to knowledge $\boldsymbol{\mathcal{E}}$. For each factor $\boldsymbol{x_k^F}$ currently observed ($\boldsymbol{x_k^F} \in \boldsymbol{\mathcal{E}}$), we compute an artificial vulnerability index $\widetilde{\boldsymbol{v}}_{\boldsymbol{i}}^{\boldsymbol{d}}\left(\boldsymbol{\mathcal{E}} \backslash \{\boldsymbol{x_k^F}\}\right)$, assuming factor k is not observed. The difference between the vulnerability indexes, $\boldsymbol{v_i^d}(\boldsymbol{\mathcal{E}}) - \widetilde{\boldsymbol{v}}_{\boldsymbol{i}}^{\boldsymbol{d}}\left(\boldsymbol{\mathcal{E}} \backslash \{\boldsymbol{x_k^F}\}\right)$, provides an indication of the influence of that factor. A similar procedure is performed for factors that are unknown to determine how informative they could be.

Results

We applied this modeling framework to the domain of elder care in collaboration with the Beijing Academy of Science and Technology. Specifically, we developed a vulnerability model to help careworkers better support aging populations, focusing on fostering independent living while at the same time identifying declining mental and physical health as early as possible. Aging populations and support for independent living are a growing concern in both developed and developing countries. In fact, the WHO estimates that the number of older people who are no longer able to look after themselves in developing countries may quadruple by 2050 [8].

To build the Bayesian network model, we have used data from the US National Institute for Health (NIH), specifically their Longitudinal Study On Aging (LSOA, description and data available at http://www.cdc.gov/nchs/lsoa/lsoa2.htm). These data are based on an initial cohort of 9,447 persons who were age 70 and over in 1995. Those persons (or their proxies when deceased) have been interviewed at three different times (waves), approximately two years apart. We have chosen to use the last two waves, as the questionnaires were more similar between those two waves than between the initial and second waves. We thus obtained access to state variables at two-year intervals. The choice of vulnerability domains and higher level vulnerability categories was mostly driven by the availability of data and by local expert knowledge [9].

Table 1 provides a list of the vulnerability domains considered, grouped into six categories. The LSOA study contains several hundred different fields, and we have selected a subset of 61 features in addition to the 34 state variables described above. Some (limited) data cleaning effort was required. We have modified fields to reduce granularity; for instance, the field *Type of Residence* contained about 15

different possible answers, which we reduced to five different types, capturing the majority of the answers. Also, fields such as *Age* or *Number of Children* were discretized into smaller numbers of states.

Table 1– Vulnerability Domains

State Variables **Category(number of variables):**list of variables
Sensation (2): Sight – Hearing
Social Involvement (2): Socially Active with friends - Socially Active with relatives
Mental Health (2): Frequency of sadness or depression, self-rated memory
Fundamental Daily Activity (6): Ability to eat, bath, dress, walk , get out of bed or chairs and use the restrooms independently
Extended Daily Activities (2) : Social Living Abilities, Cognitive Living Abilities
Medical (3) : Self-rated Health, Injured from Fall, More than three chronic conditions

The Bayesian network model was learnt using the Jsmile reasoning engine for graphical probabilistic models (available at http://genie.sis.pitt.edu/). We used the greedy-thick-thinning algorithm with various combinations of the learning algorithm parameters (type of priors and maximum number of parents allowed). While we do not have data to validate the value of the vulnerability indexes, we chose the best set of parameters for the learning algorithm by taking the one that had the best next-period prediction accuracy, which we computed by applying the Bayesian network prediction to test data that we had held from the training data. The model was then integrated and deployed into a real tool currently being deployed in a Beijing neighborhood. Figure 4 provides a snapshot of the user interface associated with a previous iteration of our model (with fewer categories). In that interface, we do not directly present the full vulnerability profile (meaning values for each of the 14 vulnerability domains) but rather aggregate each category by reporting the worst vulnerability index associated with that category (second column). In addition, we provide historical values (third column) and contributing factors (fourth column) to the care worker using the dashboard.

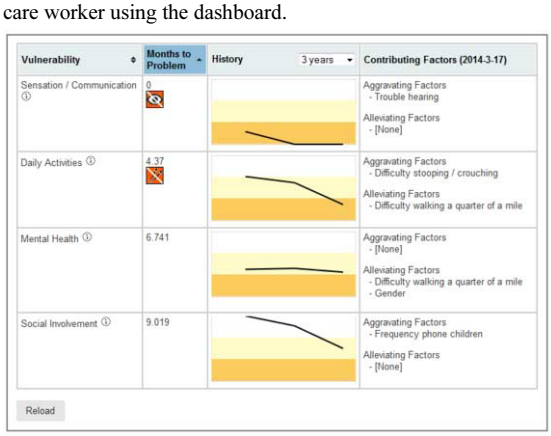


Figure 4 – Screenshot of the Eldercare Dashboard

To illustrate the effect of the known information on the vulnerability profile, we provide in Table 2 the profiles associated with four different information contents associated with a patient. Person 1 corresponds to a "Married 75 year-old man". Person 2 corresponds to the same profile, except that we also know that the person suffers from diabetes. Person 3 builds upon person 2 with the additional knowledge that the person has difficulty walking and that he feels depressed some of the time. Finally, person 4 is an even richer (and more dire) description of the person current situation. However, person 4's profile still represents incomplete knowledge of the person, capturing 10 out of the 61 factors contained in the model.

The first two columns in Table 2 are fairly similar, though there is a consistent minimal increase of vulnerability (decrease of the indexes) for person 2, who also suffers from diabetes. This is an example of cross-influences of a medical factor on social domains. The largest difference is associated with the *More Than Three Chronic Conditions* domain. As diabetes is a chronic condition, such change is logical. When comparing persons 2 and 3, we first notice that *Difficulty Walking* is now null for person 3, a direct consequence of the fact that we have observed this difficulty; it is no longer a vulnerability but rather a need. Notice again that many other vulnerability indexes have decreased as well, illustrating cross-influences. Only *Frequency Sadness and Depression* has increased, indicating lessened vulnerability. This is not surprising as feeling sad or depressed only *a little of the time* constitutes in fact positive evidence compared to the unknown state. Finally, person 4 represents a significant aggravation of person 3's state, with many additional physical and medical difficulties and a worsening of the mood. At this point, the profile enables the care coordinator to prioritize among the multiple dimensions considered, possibly leading to a focus on social living abilities and fundamental living activites as the most pressing problems for the patient, besides addressing his mental health need.

Table 2 – Illustrative Vulnerability Indexes (in Years)

	Person 1	Person 2	Person 3	Person 4
Difficulty Dressing	7.51	7.49	4.42	1.12
Difficulty Eating	15.55	15.52	11.10	5.34
Difficulty Walking	3.24	3.22	***0.00***	***0.00***
Difficulty Using Toilet	8.83	8.81	5.11	2.14
Difficulty Getting Out Chair Bed	5.77	5.74	2.97	1.10
Sight Level	6.54	6.54	5.82	4.70
Hearing Level	2.70	2.66	2.58	2.26
Socially Active Friends	12.28	12.27	11.33	9.29
Socially Active Relatives	23.12	23.11	20.58	15.98
Frequency Sadness Depression	6.65	6.61	***7.96***	**0.00**
Self-Rated Memory	32.10	32.06	31.69	31.01
Difficulty with Social Living Abilities	4.05	4.04	2.59	0.43
Difficulty with Cognitive Living Abilities	4.30	4.29	3.37	1.55
Injured From Fall	4.45	4.45	3.56	2.83
More Than Three Chronic	9.10	***7.11***	6.13	3.79
Self-Rated Health	13.92	13.78	9.69	4.50

Person 1: Married 75 year-old man
Person 2: Married 75 year old man with diabetes
Person 3: Married 75 year old man with diabetes and difficulty walking and feeling sad or depressed a little of the time
Person 4: Married 75 year old man with diabetes, hypertension, difficulty walking, stooping and finger grasping, feeling sad or depressed all of the time, who has not left his house more than 5 days in the past 2 weeks, who is a smoker and who is not able to do light house work in his home

Discussion

Person-specific standardized assessments have long been presented as one of the solutions to improve the effectiveness of social support. InterRAI [10] and the Northern Ireland Single Assessment Tool (NISAT) [11] are two suites of

coordinated assessment tools that provide such solutions in the domain of elder care. However, both NISAT and InterRAI constitute descriptive tools, seeking to capture and record an accurate profile of the patient at a given time. By contrast, the essence of our model is to predict the current and upcoming needs of a person without having to go through a complete assessment process. In that sense, it is complementary to these existing assessment suites. It can be used for triage/screening purposes to better prioritize actions to be taken, be they completing a given set of assessments, or direct action such as provision of services. Our model is also designed to bridge health and social care by providing an explicit representation of the interactions between domains.

The main impediment associated with our proposed modeling framework remains the challenge of building the underlying model from limited information. In the elder care case study that we presented, while we were satisfied with using the LSOA study, it is less than ideal as it represents an American cohort for a model to be applied in China. While some aspects of aging are universal, others are cultural and therefore local. To faciliate the construction of the model from limited data, we are currently exploring methods to build Bayesian network models from disparate, overlapping, and possibly conflicting sources of information. Indeed, in most practical cases, and especially when health and social issues are modeled together, it will not be sensible to assume that individual data will be available in one single datatable for all the variables required in the network. The novel methods that we are currently investigating will mean that the model does not require one clean dataset to be built but can be assembled from a collection of messy disparate sources of information such as

- Data sets at the individual level, covering a subset of the nodes in the network
- Aggregated information for a subset of the nodes of the network
- Expert information (influence statements and probability statements)

Specifically, we are extending methods from the artificial intelligence community. In particular, we have looked at adapting the PC algorithm [12] to build a Bayesian network from multiple overlaping datasets and from a set of expert statements extracted from the literature.

One useful feature of the model that we have not yet mentioned (and that is relevant to model building) is the fact that it can function with imprecise information. So far, we have assumed that the parameters of the model were point estimates (i.e., that we have enough data or definite domain expertise to determine a single value for each one). Again, this may not necessarily be the case in practice, either because data is limited, experts are unsure, or because multiple experts have proposed different values. In such situations, being able to handle imprecise information is essential if one does not have the time or budget to resolve conflicting information or gather more data. Both Markov chain and Bayesian network models admit imprecise extensions in the form of the Markov Set chain [13] and Credal Network [14] respectively, which enable researchers to rigoroulsy percolate the imprecision through the assessment in a logical fashion. Besides increased computational complexity, the only consequence of allowing for imprecision in the model is the fact that the vulnerability indexes would also become intervals and thus may be more difficult to interpret.

Conclusion

In this paper, we present a modeling framework for person-specific standardized vulnerability assessment in health and social care. This model combines a Markov chain model that describes the evolution of persons through states with a Bayesian network that enables researchers to customise the Markov chain parameters. The purpose of the model is to estimate the vulnerability profile of a person based on limited input information so as to provide health professionals and social care workers with a holistic view of the person's profile when they receive new situations or update existing ones. We present an application in the context of elder care though our model that can be applied in a variety of health- and social-care scenarios and is especially relevant for cases that bridge multiple domains.

Acknowledgments

This project was supported in part through a grant from the Irish Innovation Development Agency (ref. 133954,September 2009).

References

[1] OECD. Social spending after the crisis. OECD, 2012

[2] Truffer CJ, Klemm JD, Wolfe CJ, Rennie KE, and Shuff JF. Actuarial Report on the Financial Outlook for Medicaid. US department of Health & Human Services. 2012

[3] Mann C. Medicaid and CHIP: On the Road to Reform. Presentation to the Alliance for Health Reform/Kaiser Family Foundation March 4,2011

[4] Gawande A. The Hot Spotter. Can we lower medical costs by giving the neediest patients better care? The New Yorker, January 24, 2011

[5] Tarlov AR. Public Policy Frameworks for Improving Population Health. Annals of the New York Academy of Sciences 1999: 896: 281-293.

[6] Cutler DM and Lleras-Muney A. Education and health: evaluating theories and evidence. No. w12352. National Bureau of Economic Research, 2006.

[7] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988

[8] WHO. 10 facts on ageing and the life course. WHO, 2012 Available at: http://www.who.int/features/factfiles/ageing/en/index.html (Accessed 2 April 2013)

[9] Ability assessment for older adults. ICS 03 080 01. 2013

[10] Gray LC, Berg K, Fries BE, Henrard JC, Hirdes JP, Steel K, and Morris JN. Sharing clinical information across care settings: the birth of an integrated assessment system. BMC health services research,2009: 9(1):71.

[11] Taylor BJ. Developing an integrated assessment tool for the health and social care of older people. British journal of social work, 2012:42(7):1293-1314.

[12] Spirtes P , Glymour, CN, and Scheines R. Causation, prediction, and search. Vol. 81. MIT press, 2000.

[13] Hartfiel DJ. Markov Set-Chains. Series: Lecture Notes in Mathematics, Vol. 1695, 1998

[14] De Campos CP and Cozman FG. The inferential complexity of Bayesian and credal networks. *IJCAI*. 2005:5.

Address for correspondence

Léa A.Deleris. RSM and Manager at IBM Research – Ireland.
Lea.deleris@ie.ibm.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-467

Trust, Perceived Risk, Perceived Ease of Use and Perceived Usefulness as Factors Related to mHealth Technology Use

Rebecca Schnall [a], Tracy Higgins [a], William Brown [b, c], Alex Carballo-Dieguez [c], Suzanne Bakken [a,b]

[a] *Columbia University School of Nursing, New York, NY, USA*
[b] *Columbia University Department of Biomedical Informatics, New York, NY, USA*
[c] *HIV center, Division of Gender, Sexuality and Health NYS Psychiatric Institute and Columbia University New York, NY, USA*

Abstract

Mobile technology use is nearly ubiquitous which affords the opportunity for using these technologies for modifying health related behaviors. At the same time, use of mobile health (mHealth) technology raises privacy and security concerns of consumers. The goal of this analysis was to understand the perceived ease of use, usefulness, risk and trust that contribute to behavioral intention to use a mobile application for meeting the healthcare needs of persons living with HIV (PLWH). To understand these issues, we conducted focus group sessions with 50 persons living with HIV and 30 HIV healthcare providers. We used the e-commerce acceptance model to analyze our focus group data. Findings from the study demonstrated the need for mHealth to be perceived as useful, easy to use, with little perceived risk accompanied by a measure of trust in the creators of the technology. Findings from this work can inform future work on patients and providers' perceptions of risk, trust, ease of use and usefulness of mHealth technology.

Keywords:

mobile health; HIV; privacy; trust; security

Introduction

Mobile technology is nearly ubiquitous as increasing numbers of people carry a mobile device [1]. As such, mobile technology offers a unique opportunity for the diffusion of behavioral health interventions. Mobile health (mHealth) technology can guide decisions about health by providing round-the-clock real-time feedback [2] and opportunities for consumers to learn about and manage their disease.

These characteristics of timeliness and accessibility are particularly relevant to care for HIV which has shifted from a disease with a high mortality rate to one requiring lifetime chronic care and health monitoring. mHealth tools can help address many of the healthcare needs of persons living with HIV (PLWH) to promote retention in HIV care and adherence to medication regimens[3]. In the US, 1.2 million people are currently living with HIV [4], with worldwide estimates at 35 million [5]. A number of studies have shown promise for the use of mHealth for supporting HIV care [6-9].

Our study population included PLWH who were the targeted end users of a smartphone application (app) designed for monitoring and managing their health and for communicating with their care providers. PLWH is a relevant study population due to both their need for self-management support in chronic care as well as the persistent cultural stigma associated with HIV/AIDS [10].

Despite the promise of mHealth in monitoring health behaviors and improving the delivery of health care, little is known about users' concerns over privacy and information sharing using mHealth technology. Privacy has been a topic of interest to researchers in psychology, sociology and more recently information technology (IT) [11-13]. Existing research on privacy emphasizes data security and confidentiality, largely focusing on electronic medical records. Yet consumers concerns around privacy related to mHealth technology use remain poorly understood. There are a number of issues related to privacy and information security in HIT that need to be carefully monitored and addressed in order to offer a technology most likely to engage consumers. The purpose of this study was to understand the perceived ease of use, usefulness, risk and trust that contribute to behavioral intention to use a mobile app for meeting the healthcare needs of PLWH.

Theoretical Framework

We used the e-commerce acceptance model (EAM) to understand technology acceptance of mHealth for PLWH [14]. The EAM has not been previously studied in the context of health or mobile technologies. It builds on the most widely employed model of IT adoption, the technology acceptance model (TAM) which we have previously studied in the context of the delivery of HIV care through an electronic record [15].

The dependent variable in this model is the same as in the TAM: behavioral intention to use the technology. The key constructs in the TAM are perceived usefulness and perceived ease of use [16]. In the case of health information technology (HIT), perceived usefulness is defined as the extent to which a person believes that using the HIT system will enhance his/her outcomes. Perceived ease of use is defined as the degree to which system use will be free from additional effort [17].

In the EAM, trust and perceived risk are two key drivers added to the TAM variables using the overall structure of the Theory of Reasoned Action (TRA) [18]. Integration of the four independent variables were tested and their relationships were validated. The model proposes that these four drivers lead to the primary constructs of "intention to transact" and "on-line transaction behavior" in the e-commerce sector, which in many ways may mimic consumer behavior in eHealth.

Trust is a behavioral belief that has been studied in e-commerce and has been shown to have a favorable effect on consumers' intent to use a technology. Trust may be defined as the belief that the other party will behave responsibly and will not attempt to exploit the vulnerabilities of the user [14]. There are two kinds of trust: 1) party trust - trust in the

benevolence and credibility of a particular party; and 2) control trust - trust in the integrity of the transaction medium. The importance of trust is heightened when there is a high degree of uncertainty which can happen in the case of HIT when users don't understand where their information is being stored or how it is being transmitted.

For mHealth use, we would expect that consumer trust in the mHealth app is likely to influence behavioral intention to use the technology [19]. Trust is also related to perceived usefulness because users may not be certain that outcomes may be achieved unless there is confidence in the entities behind the HIT system. Trust is related to perceived ease of use because it reduces the efforts that would otherwise be necessary to monitor the proper functioning of the system. Trust is also related to perceived risk as a higher level of trust in an on-line entity reduces the perception of risk during the interaction.

Perceived risk is an important factor for eHealth use in light of the distance, in both time and space, between the consumer and the healthcare entity, as with most on-line transactions. It may be described as a degree of uncertainty related to use of the medium that is beyond the control of the information manager associated with the eHealth service. The consumer risks the possibility of suffering a loss while using the technology [20]. mHealth users perceive risk when the security of the infrastructure for securing their personal health information is not verified. Consumer perception of risk is negatively associated with intention to transact. Similarly, if an end-user perceives the risk of disclosure of their personal health information to be low then they are more likely to use a mHealth app.

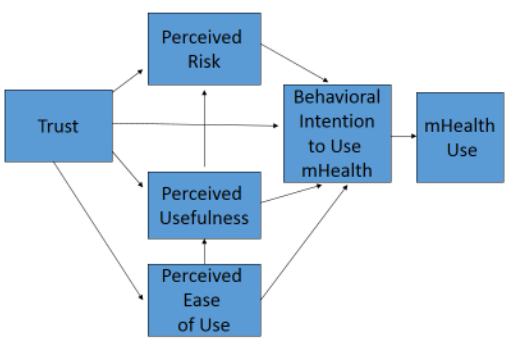


Figure 1– E-Commerce Acceptance Model applied to mHealth Technology Use

To better understand the applicability of this model for mHealth technnology, we used focus group data which was collected as part of a larger study to to inform the development of a mobile app for HIV treatment and care for PLWH. The purpose of the larger study was to elicit from potential app users and their care providers the desired content and features of a mobile app for meeting and improving the health care needs of PLWH. The goal of this analysis was to understand the EAM constructs that contribute to behavioral intention to use a mobile app for PLWH.

Methods

Design

This study applied user-centered participatory design methods to inform the content of a mobile app for HIV treatment and care. The qualitative study was implemented using focus group methodology.

Sample

The sample comprised PLWH and clinicians and case managers who provide care to PLWH.

Procedures

The Columbia University Medical Center Institutional Review Board reviewed and approved this study protocol. Focus group sessions were moderated by a trained facilitator using a focus group guide. Each focus group session was also attended by at least one additional study team member who recorded notes and tended to logistics such as consent forms, food, and compensation to ensure that the groups ran smoothly. Following completion of the individual informed consent process and prior to the start of each focus group session, participants completed a demographic survey.

Each PLWH focus group session took place in a conference room at the Columbia University School of Nursing campus, lasted between 48-72 minutes and was conducted in English or Spanish. The HIV clinicians and case managers' focus groups took place at a conference room at their workplace in the following locations: New York Presbyterian Hospital, the Brooklyn Hospital Path Center, and AIDS Service Center, a community-based organization. We gave $25 to our PLWH participants and $50 to the clinicians and case managers as a token of appreciation for their time. Food appropriate for the time of day was served during each focus group session. All focus group sessions were audio-recorded and transcribed for analysis.

We asked a series of questions during our focus group sessions to identify the functional specifications of a mobile app for PLWH. At the end of our focus group sessions, we specifically asked focus group participants: "What are some of the privacy and confidentiality concerns that you anticipate encountering when using a mobile health app on your phone?"

Data collection continued until saturation was reached, which occurred when we heard similar patterns and themes across groups and no new information was being shared [21].

Data Analysis

Descriptive statistics of the demographic information were calculated using SPSS version 21. To support the credibility of the data, we used "member checks," i.e., sharing of initial data interpretations with participants to ensure accurate interpretations. In addition, after each focus group, we conducted peer debriefing and triangulated findings across participants and facilitators. Research team members subsequently coded the transcripts using NVivo™ (QSR International, Victoria, Australia) software. We coded the data using the four constructs of the E-commerce acceptance model to deductively code focus groups to understand factors related to mHealth technology use. Coding of transcripts continued until consensus was reached.

Results

Sample

We conducted six focus group sessions with PLWH (n=50). Two of the focus groups were conducted in Spanish and one of the groups was comprised of only female participants. There were 37 male participants and 13 female. Participants' ages ranged from 18-59. Participants self – identified as Black (N =26), White (N =7), Pacific Islander (N = 1), and American Indian (N =1) and the remainder of the participants

did not report their race. Half of our participants (N =25) self-identified as being of Hispanic/Latino ethnicity.

In addition, we conducted three focus group sessions with HIV care providers which included clinicians and case managers. All focus groups were conducted in English. Participants' ages ranged from 23-62 years and included 5 males, 24 females and 1 transgender female. Participants identified as Black (N =9), White (N =17), Pacific Islander (N = 1), and Asian (N =3). Six participants self-identified as being of Hispanic/Latino ethnicity.

Findings

We used the four constructs of the EAM to understand factors related to mHealth use in PLWH. Our focus group findings were organized according to the four broad constructs: trust, perceived risk, perceived usefulness, and perceived ease of use (Table 1).

Table 1– EAM Constructs and Sample Quotes

Construct	Sample Quotes
Trust	• I am wondering why they're interested in all of this (creating and HIV app) all of a sudden." [PLWH]
Perceived Risk	• One of barriers is a concern that you'll be able to track the person. [PLWH]
Perceived Usefulness	• It's very non-confrontational that way, and I think that's a big benefit of texting [Healthcare Provider]. • It would be really good if the doctor or CDC put that in an app so that you have access to it and you can show it right then and there [PLWH]
Perceived Ease of Use	• Just to learn a new thing (app/ smartphone use) it's becoming difficult [PLWH]

Trust

PLWH recognized the need for security safeguards and controls so that they could trust using an mHealth app. For example, one participant said, "provided the company that furnishes the app, has the right security so the information does not leak. Having that security, then there is no problem." Another participant cited an example of a security system tied to the device which would make him feel comfortable using a mHealth app for PLWH. "The warranty from the company that makes it…the security the company has of itself, which in this case, well, benefits us, so that information does not leak." Participants had concerns over who would have access to their information and didn't trust certain entities to access their personal health information.

Perceived Risk

A number of participants expressed concerns over the risk of who would have access to their information, where it would be stored and if their HIV status had the potential to be disclosed. One participants said, "It's like she said about technology, you don't know where this information is going." Another participant had the perception that anyone can access any information that is on the Internet. She said "Because right now anybody could get into the Internet and go into your information."A younger participant in another focus group which was comprised of 21-24 year olds said, "As far as using an app I would probably never put information in, I wouldn't personally because technology is out of order now. People can find out anything. You can do the fingerprint thing. Where is this information going to, who is seeing it, because it's being documented? Where is it going?" Since smartphones have the functionality to detect your location, participants perceived a great risk to being tracked. For example one participant said "It's like I don't know…with the location settings with the iPhone and all that stuff." Participants did acknowledge that a certain amount of risk may be acceptable or unavoidable. One focus group participant said, "We just have to get, how can I say it, a little secure about it. Because I didn't like, where I worked at, I didn't like them scanning all my information and putting it on the computer. Because the computer can be hacked in, in any given time, you understand? But I got comfortable with it. So, it's a period you got to go through. You've got to adjust to certain things in life."

Perceived Usefulness

PLWH thought that mHealth had the potential to be very useful for their care. For example one patient thought that it would provide updates on medications. "Like new medications. Every couple of months they come out with new HIV medication." Another PLWH thought the mHealth would be a useful tool for creating reminders "I think some phones have a memo device where you could actually talk into it and say…" Participants also thought it would be useful for learning more about their disease or sharing information with others. For example, one PLWH said, "Since we already have our immune system compromised, we should know for example if we have something, know what it is, and the faster the better so we take care of our health."

Providers, like patients, thought that mHealth can be very useful in the delivery of care to PLWH in a private manner. For example, one provider said, "Think about it, when they're on the phone somebody else can hear their end of the conversation. If they're texting, no one knows what's going on." Providers also thought that mobile technology was important for their patients to be able to self-manage their care. "Some of them use their phones for alarms. They have alarms set on their phones and they have calendar applications where they can write appointments."

Perceived Ease of Use

Both PLWH and providers described a number of factors related to ease of use that would facilitate use if present and otherwise would present a barrier, such as reliable phones, simplicity, and operation autonomous from an Internet connection. One PLWH said that "I would be willing to use the app if I have the phone with no problems." Another participant echoed a similar sentiment and said that "because sometimes there are some apps that are a bit complicated and well people get frustrated immediately, and later they don't use them. It has to be something simple." Both PLWH and providers thought that an app should not rely on Internet connectivity. Specifically, one provider said "An app would be helpful if certain parts of the app didn't need internet."

There were also a number of patients and providers who described the need for a simple, easy to understand app. A provider commented on the need for information presentation to be understandable, even to patients with "cognitive disability." Specifically, she said, "So even a graph that can be simple to us may not be something they're easily able to read." Another healthcare provider, in a different group echoed a similar idea and said, "It has to be simple enough that the person can really get hooked to it."

Discussion

The increase in the use of mobile technology has the potential to improve the delivery of healthcare, but it has also raised new concerns over privacy and security of personal health

information [22]. At the same time little attention has been paid to understanding technology acceptance in mHealth with a particular focus on privacy and security. Mobile devices and apps have created a new set of privacy concerns since protected health information can be stored on mobile devices, processed within these apps or shared via networks that are not secure [23]. While mHealth holds great promise, if patients are not willing to use the technology because it is not perceived as useful, easy to use or they have concerns over data security then it will not be able to improve health outcomes.

Concerns over mHealth acceptance and use becomes particularly relevant as patients and consumers begin to transfer their personal health information from their mobile devices to a central electronic health record. For example, Apple has started a partnership with Epic [24]. In addition, open platforms such as Apple's Healthkit [25] and Google Fit allow patients to transfer their personal health information to a central electronic health record, yet these systems are still quite limited and are focused on fitness.

In this study, we sought to examine PLWH and their healthcare providers' perceptions of the technology acceptance factors necessary for PLWH to use an HIV app to manage their health. Many study participants thought that a mobile app for PLWH would be extremely useful.

Participants also stressed the need for the app to be easy to use, consistent with past research in HIV care. In particular, participants suggested the app not rely on Internet connectivity. Both PLWH and providers described the cognitive deficits that are often experienced by PLWH and therefore the need for the app to be very simple and straightforward.

At the same time, there was an underlying concern over security of their information. Participants throughout our sessions wanted to understand who would have access to their personal health information. They emphasized the need to trust the "owner" of the app and if a particular company or institution was associated with the app then they would not be willing to use the app for accessing, entering or sharing their personal health information.

Perceived risk was an underlying theme throughout our nine focus group sessions. Participants were concerned over where their data was being stored and who would have access to their data. Their concerns included trust in the institutions that collect, store, or transmit their data as well as general apprehension about data transfer using the Internet network and technological devices which are continually under development. In the US, there have been a number of policy initiatives at the federal level to protect patient health data being transferred electronically. Of note, the Markle Common Framework has a set of published principles that provide the foundation for managing personal health information within consumer accessible data streams[26]. The Markle Connecting for Health Framework has guided the US government implementation of IT including the Health Information Technology for Economic and Clinical Health (HITECH) Act [27].

Even with this guiding set of principles which is quite extensive and comprehensive, consumers remain concerned that their personal health information is not secure and they have particular concerns regarding the risks associated with mobile technology in sharing and accessing their personal health information.

We did not specifically ask study participants about their intention to use this technology. Instead, we examined study participants' perceived ease of use, usefulness, trust and privacy. The lessons learned from our examination of each of these constructs are important for the development of this technology especially in a highly stigmatized population [28]. As a result, the intersection between the findings from each of our study constructs is likely to predict a PLWH's behavioral intention to use a mobile app for managing his/ her health.

Limitations

We conducted this study in New York City, which limits the generalizability of the findings to one geographic area. Even so, New York City has the greatest number of PLWH in the Unites States making this an appropriate setting for conducting this work [29]. In addition there may have been a response bias from our study participants since our methods relied on self-report.

Conclusion

Findings from this work can inform future work on patients and providers' perceptions of risk, trust, ease of use and usefulness of mHealth technology. Findings from this work have broader applicability to others living with chronic diseases as well as any individuals who seek to use mobile technologies for accessing, storing and sharing their personal health information. Future development of mHealth technology needs to integrate ease of use, usefulness, trust and perceived risk to facilitate the use of mHealth technology for consumers.

Acknowledgments

This publication was supported by a cooperative agreement between Columbia University School of Nursing and the Centers for Disease Control and Prevention (CDC; 1U01PS00371501; PI: R Schnall). The findings and conclusions in this report do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

[1] Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science. 2010;330(6005):759-60.

[2] Kumar S, Nilsen W, Pavel M, Srivastava M. Mobile Health: Revolutionizing Healthcare Through Trans-disciplinary Research. Computer. 2013;46(1):28-35.

[3] Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth Technology as a Persuasive Tool for Treatment, Care and Management of Persons Living with HIV. AIDS Behav. 2015:1-9.

[4] Centers for Disease Control and Prevention. HIV in the United States: At a Glance 2014 [cited 2014 November 30]. Available from: http://www.cdc.gov/hiv/statistics/basics/ataglance.html.

[5] World Health Organization (WHO). World AIDS Day 2014 - Closing the gap in HIV prevention and treatment services 2014 [cited 2014 November 30]. Available from: http://www.who.int/hiv/en/.

[6] Catalani C, Philbrick, W., Hamish, F., Mechael, P., Israelski, D. mHealth for HIV Treatment & Prevention: A Systematic Review of the Literature. The Open AIDS Journal. 2013;7:17-41.

[7] Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. The

Cochrane database of systematic reviews. 2012;3:CD009756.

[8] Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The Effectiveness of Mobile-Health Technology-Based Health Behaviour Change or Disease Management Interventions for Health Care Consumers: A Systematic Review. PLoS Med. 2013;10(1):e1001362.

[9] Pellowski J, Kalichman S. Recent Advances (2011-2012) in Technology-Delivered Interventions for People Living with HIV. Curr HIV/AIDS Rep. 2012;9(4):326-34.

[10] Mahajan AP, Sayles JN, Patel VA, Remien RH, Sawires SR, Ortiz DJ, et al. Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. Aids. 2008;22 Suppl 2:S67-79.

[11] Culnan MJ, Armstrong PK. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. Organization Science. 1999;10(1):104-15.

[12] Stewart KA, Segars AH. An Empirical Examination of the Concern for Information Privacy Instrument. Info Sys Research. 2002;13(1):36-49.

[13] Dinev T, Hart P. Internet privacy concerns and their antecedents - measurement validity and a regression model. Behaviour & Information Technology. 2004;23(6):413-22.

[14] Pavlou PA. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. International Journal of Electronic Commerce. 2003;7(3):101-34.

[15] Schnall R, Bakken S. Testing the Technology Acceptance Model: HIV case managers' intention to use a continuity of care record with context-specific links. Informatics for health & social care. 2011;36(3):161-72.

[16] Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 1989;13(3):319-40.

[17] Adams DA, Nelson RR, Todd PA. Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication. MIS Quarterly. 1992;16(2):227-47.

[18] Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.

[19] Jarvenpaa SL, Tractinsky N, Saarinen L. Consumer Trust in an Internet Store: A Cross-Cultural Validation. Journal of Computer-Mediated Communication. 1999;5(2):0-.

[20] Bauer RA. Consumer behavior as risk taking. . Cox DF, editor. Cambridge: Harvard University Press; 1960.

[21] Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage; 1998.

[22] Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile Health Technology Evaluation: The mHealth Evidence Workshop. American Journal of Preventive Medicine. 2013;45(2):228-36.

[23] Luxton DD, Kayl RA, Mishkind MC. mHealth data security: the need for HIPAA-compliant standardization. Telemedicine journal and e-health : the official journal of the American Telemedicine Association. 2012;18(4):284-8.

[24] Carr D. Apple Partners With Epic, Mayo Clinic For HealthKit 2014. Available from: http://www.informationweek.com/healthcare/mobile-and-wireless/apple-partners-with-epic-mayo-clinic-for-healthkit/d/d-id/1269371.

[25] Dolan B. Why Apple's HealthKit launch is important for digital health 2014 [cited 2014 September 6]. Available from: http://mobihealthnews.com/33765/why-apples-healthkit-launch-is-important-for-digital-health/.

[26] Connecting Consumers: Overview and Principles New York: Markle Foundation; 2011 [cited 2014 December 6]. Available from: http://www.markle.org/health/markle-common-framework/connecting-consumers/overview.

[27] Blumenthal D. Launching HITECH. New England Journal of Medicine. 2010;362(5):382-5.

[28] Shacham E, Rosenburg N, Önen NF, Donovan MF, Turner Overton E. Persistent HIV-related stigma among an outpatient US clinic population. International Journal of STD & AIDS. 2015;26(4):243-50.

[29] HIV Epidemiology and Field Services Program. HIV Surveillance Annual Report. New York, NY: 2013.

Address for correspondence

Rebecca Schnall, RN, MPH, PhD
Assistant Professor of Nursing
Columbia University
617 W. 168th Street
New York, NY 10032
Phone: 212-342-6886

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-472

Electronic Health Record System Contingency Plan Coordination: A Strategy for Continuity of Care Considering Users' Needs

Marcela T. Fernández, Adrián R. Gómez, Américo M. Santojanni, Alfredo H. Cancio, Daniel R. Luna, Sonia E. Benítez

Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina

Abstract

Electronic Health Record system downtimes may have a great impact on patient care continuity. This paper describes the analysis and actions taken to redesign the Contingency Plan Procedure for the Electronic Health Record System of Hospital Italiano de Buenos Aires.

After conducting a thorough analysis of the data gathered at post-contingency meetings, weaknesses were identified in the procedure; thus, strategic actions were recommended to redesign the Contingency Plan to secure an effective communications channel, as well as a formal structure for functions that may support the decision-making process. The main actions were: 1) to incorporate the IT Contingencies Committee (Plan management); 2) to incorporate the Coordinator (general supervision of the procedure); and 3) to redefine the role of the Clinical Informatics Resident, who will be responsible for managing communication between the technical team and Electronic Health Record users. As users need the information for continuity of care, key users evaluated the impact of the new strategy with an adapted survey.

Keywords:

Downtime; IT Management; Survey; SWOT Analysis.

Introduction

Continuity of patient care as a health care process acquires great potential through information technology (IT). However, when applications or systems fail, the technologies become a dangerous vehicle for the perpetuation of erroneous information that may lead to mistakes in diagnosis and higher monetary costs [1]. The critical nature of hospital care impacts patient information management, therefore institutions are forced to plan and formalize courses of action to maintain their IT infrastructure at times of crisis[1].

During the review of the events experienced at Hospital Italiano de Buenos Aires (HIBA) and relevant literature [2,3], it has been noted that when all instances of "redundancy" and "control" designed to support and guarantee service continuity are exhausted, the institution's own alternative procedures are implemented in order to protect all crucial information. The impact of technology and informatization on hospitals' operating processes must be addressed as a complex process developed over time where an institution-wide transformation process is evident [3]. This process involves knowing and rating systems' "life cycles", as well as the importance of weighting the lessons learned from disasters and crises that may jeopardize patients' data management. In order to offer continuity of patient care, technology is useful if the regulation, standardization, and protection of critical operating processes and relevant information are solved beforehand [4]. Therefore, it is essential to make periodic adjustments to the action plans designed by the institution for contingencies. The objectives of this study are to describe the redesign of the contingency plan of the Electronic Health Record (EHR), as well as to know the users' perceptions of it.

[1] Crisis or disaster recovery, such as power failures, general hardware failures, logical or physical breaks in data connections, or downtimes planned for application updates.

Materials and Methods

Background

The HIBA is a high complexity teaching hospital founded in 1853. It is member of a nonprofit health care network that runs a second hospital, 25 outpatient clinics, and 150 doctor's offices distributed throughout the city of Buenos Aires and the surrounding metropolitan area. This infrastructure also includes 750 hospital beds, 200 of which are prepared for intensive care, 1200 home care beds, and 41 operating rooms. The network is composed of 2800 physicians, 2800 health care agents, and 1900 management and administration staff.

A Health Information System has been gradually implemented since 1998, based on an in-house development process where medical and administrative information is managed from data gathering to data analysis. A modular, patient-centered, web-based EHR focused on problems, and known as ITALICA, has been implemented by the hospital. This EHR may be used to register outpatient care, hospitalization, emergencies and home care related data.

There is a formal procedure that defines the actions and forms of communication for maintenance of the processes while the EHR system recovers from a significant disruption.

This procedure was implemented in planned and unplanned downtimes. The procedure was tested during the following year, and after every downtime there were meetings with key users of the EHR system in order to learn of the impact to the workflow. Since the preliminary results showed inefficient communication, the procedure was redesigned.

The key users involved the most critical processes of the hospital, such as the Emergency Department, Laboratory, Nursing Department, Patient Admission area, Imaging Department; and specialized areas of the Health Informatics Department (HID) such as the Help Desk, Software Engineering, Clinical Informatics, and Policies & Procedures involved in the development process and operations for the activation and deactivation of the Plan. The IT Contingencies team is composed of key users of the EHR and members of the HID staff.

Definitions

For the purpose of this research, the following definitions were used (see Figure 1):

- Patient Care Continuity [5]: Different aspects referred to the treatment to be applied to a specific health condition, from both chronological and geographical standpoints.
- IT Contingency Plan: A program detailing the methods and means necessary to control risk in information systems and to minimize negative consequences.

In this paper, we have analyzed the "high availability" and "IT Contingency Plan" concepts as two different yet supplementary notions. The "high availability" and "IT Contingency Plan" principles are IT management tools that represent the purpose of "continuity". The first concept refers to technology and information systems-based mechanisms, while the second one secures continuity based on alternative procedures that encompass all of the hospital's resources, including paper records of patients' medical evolution.

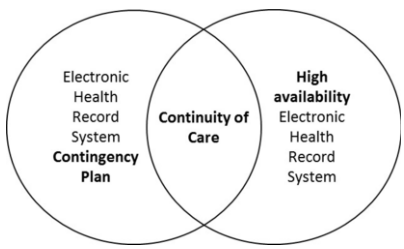


Figure 1 - Institutional Framework for Continuity Care

The redesign process is diagrammed in the following figure (see Figure 2). The first step was the SWOT analysis of the original procedure. A survey administered to key users was the next step. Finally, a new version of the survey was designed and validated by a committee of experts.

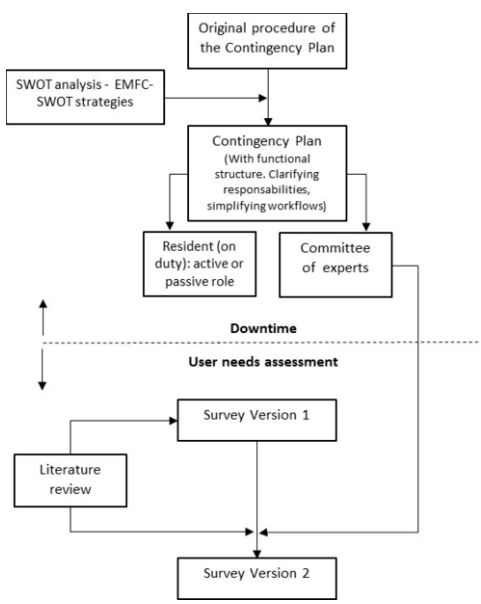


Figure 2 - Global Chart of the Redesign Process

Redesign

A SWOT analysis (see Table 1: strengths, weaknesses, opportunities and threats) was performed to redesign the procedure, and four possible implementation strategies were prepared[2]. Additionally, the work structure applied was based on rules and standards.

On the one hand, the NIST special publication 800-34 [6] and the guidelines for Contingency Plans preparation and drafting were considered to structure the Contingency Plan management and coordination; and on the other hand, the ISO 22301:2012 standard about "business continuity" management was considered to approach the "Plan-Do-Check-Act" [7] process applied to plans and procedures maintenance. Even though COBIT Management Guidelines [8] were considered, this paper is focused on users' needs.

Table 1 – SWOT Analysis of the Procedure of the Contingency Plan of EHR N°98/12 01 00

Strengths	Opportunities	Weaknesses	Threats
Technical capabilities of key areas of the HID	The Institution has specialized technicians and analysts to define processes and procedures	Lack of a clear position that handles the coordination and supervision to ensure the effective conduction of the proceedings	Facing crisis and contingencies, regarding critical services, is a process that relies on "key actors"
Includes a flow of communication for the activation and deactivation of the plan			
Use of software especially designed for contingencies	There is an institutional interest in reviewing plans based on the Policy "Continuity of Patient Care"	The procedure should clarify the active or passive mode of the workflow by the Resident on duty	The level of participation of critical sectors in post-contingency meetings does not meet the expectations
Clearly defined times of response			

EMFC-SWOT [9] Matrix results (Explore, Maintain, to Face, Correct) were obtained as a result of the SWOT analysis performed (see Table 2).

Impact assessment

With an adapted survey [10], key users assessed the impact of the new strategy. The redesigned procedure was evaluated after the unplanned downtime that took place on June 24, 2014.

With the findings of the first adapted survey and the literature review [11], a new version of the survey was designed. Once the preliminary list of questions was elaborated, the steps to follow were: (a) assessment by a committee of experts, (b) pretesting, (c) the validation itself, and (d) implementation. Pretesting was conducted with key users, who invited new users to participate in successive iterations.

[2] The analysis universe was the EHR Contingency Plan Procedure N° 98/12 01 00. Strengths and weaknesses refer to the features of the procedure (internal environment), while opportunities and threats refer to the microenvironment context, in this case the HIBA.

The EHR system was affected by the unplanned downtime that took place on June 24, 2014. The post-contingency review was conducted on June 25 and July 17, 2014.

Results

Tests of the original procedure were conducted from December 2012 (formal publication) to June 2014. The post-contingency documentation and data analysis were conducted between April and June 2014.

As a major input, the SWOT analysis suggested a new allocation of responsibilities regarding procedural coordination and supervision. Strategies derived from the analysis were implemented and continue to be in place to-date (EHR Contingency Plan Procedure N° 98/12 01 01), taking into account the following strategic EMFC-SWOT Matrix that considers the four most relevant strategies to be followed:

Table 2 – EMFC-SWOT Matrix. Strategies to correct weaknesses detected in the procedure

Evaluation		Strategy		Action
weaknesses	Lack of definition in the coordination and monitoring of the process	**correct**	Develop a well-defined functions' framework	**I**
threats	Strong dependency on "key actors" on critical services	**to face**	Develop a procedure for activation and deactivation of the Plan, and update the instruction manuals in critical areas	**I**
strengths	Use of software especially designed for contingencies	**maintain**	Expand the use of contingency focused applications to all critical services	**V**
opportunities	There is a Policy for the Patient Care Continuity Plan	**explore**	Develop a Standard / Policy on the IT Contingency Plan	**P**

***I**: Immediate application; **P**: Programmed; **V**: Viability should be studied*

New procedure

The following recommendations are made based on the findings of the procedure's post-contingency analysis:

- Incorporating two new players: the *IT Contingencies Committee*[3] (Plan management), and the Coordinator (general supervision of the procedure).
- Redefining the role of the *Clinical Informatics Resident,* who will be responsible for managing communication between the technical team and EHR users.

[3] This is a department Committee formalized by Policy N° 73/14 IT Contingency Plan. The IT Contingencies Committee is made up of the Head of HID, the Deputy Head of HID, the Head of Clinical Informatics, the Head of IT, and the Head of Software Engineering.

Findings of the first version of the survey

The redesigned process was subjected to evaluation through a survey conducted post-contingency of the unplanned downtime of June 24, 2014. The survey showed:

Table 3 – Survey Results. Unplanned downtime, included EHR: June 24, 2014. Post-Contingency survey: June 25 & July 17, 2014

Questions	A	B	C	D	E
1. How downtime affected the workflow?	Great Impact	Moderate	No impact	No impact	No impact
2. Were there problems to communicate with HID?	YES	NO	NO	NO	NO
3. How long did you wait to launch the Contingency Plan?	10'	Up to 5'	Up to 5'	1 hour	1 hour

***A**: Imaging services Department; **B**: Laboratory; **C**: Emergency Department; **D**: Nursing Area; **E**: Admissions*

Validation of the second version of the survey

The operationalization of variables and the questionnaire design were validated by the IT Contingencies Committee for the purpose of measuring users' opinion on the redesigned procedure and how it affected their workflow from downtime to uptime.

The new survey variables are: knowledge of the procedure; Plan implementation; user's response capacity; accessibility during downtime; post-contingency data restoration; communication (between the HID and the areas); workflow impact; registration possibilities; and the time needed to activate the Plan.

The pre-testing, started on January 19, 2015, assessed the post-contingency of the downtime planned for January 7, 2015. This assessment was conducted by 12 representatives of the hospital's critical areas or services. The pretesting resulted in the reorganization of questions and the modification of some words for a better understanding.

The most important pretesting findings were related to communication, an alternative work plan, and training. Users perceived that communication with the HID has improved. As for planned downtime, they have recognized to have an alternative work plan when downtime occurs. As for unplanned downtime, they stated that training is necessary.

Discussion

This paper describes the analysis and actions taken to redesign the HIBA's EHR Contingency Plan. It also provides answers to questions regarding the importance of regularly reviewing plans, as well as of formally describing a functions structure.

The procedure's redesign was changed while structuring functions, and it triggered the creation of an IT Contingencies Committee that comes into play in case of contingencies so as to facilitate the Plan's management coordination, supervision, and communication. The original procedure showed the flow of activities were conducted so that the HID may activate and

deactivate the Plan, but it did not clearly describe the decision making process. Restructuring functions and responsibilities while redesigning the Plan turned out to be essential to providing the team with adequate leadership, and to make decisions despite the uncertainties proper of any contingency.

According to the findings of the first version of the survey that was performed after the unplanned contingency, the workflow of the imaging services department was the most affected, as, unlike all other services, communications with the HID reported to be delayed. During the downtime period, most services used paper-based forms while only two services used the software and hardware specially designed to face the contingency.

Although it may be assumed that recommendations made regarding the IT Contingencies Committee, the procedure's supervision, and the coordination role to improve communication with all other sectors, will positively affect the Plan's execution, these results may not be generalized, as the survey was only conducted on key users. In addition, the survey (see Table 3) was translated from English to Spanish and it did not take into account cultural differences, nor was it validated. The second version of the survey, although already validated by the Committee, is still being tested. The main findings of this study are from first survey and the pretesting step of the second survey. The final version is expected to be conducted on general users by June 2015, after correction and validation.

As for the implications of the results, the need to regularly review procedures that are part of contingency plans becomes apparent in the systems' "lifecycles". These are not static procedures; they keep pace with the dynamics of IT systems that may change either gradually or abruptly as cutting edge technology moves forward. Therefore, the maintenance stage is an important part of the EHR Contingency Plan.

An important aspect of this paper is how users' needs have been approached. Despite the fact that authors do not pay much attention to this [12,13], it is important to consider users because they are in contact with the information system and are responsible for recording the patient's clinical evolution.

Based on our findings, we suggest three future lines of work:

1. Performing a new IT risk analysis to supplement the Standard on IT Contingency Plan and the institutional Standard/Policy on Continuity Plan;
2. Conducting the EHR downtime survey validated by the IT Contingencies Committee;
3. Developing contingency applications for internal support purposes (redundancy of elements).

Because of the importance of EHR systems for hospitals [14], regularly reviewing the Contingency Plan designed for such system is paramount to achieve care continuity [15]. This review must include an efficient communication channel as well as a formal function structure that may support the decision making process.

Conclusion

First, it is important to recognize the dynamic aspect of organizations, which involves regularly reviewing their contingency plans; second, in order to make efficient decisions in a contingency scenario, it is important to have a structure in place that may guarantee the Plan's direction and supervision. As for the approach, taking into account users is key to any Contingency Plan; as in any ideal situation, it is they who should guarantee that all the information needed to achieve care continuity will be recorded.

Acknowledgments

The authors gratefully acknowledge the contribution of the "IT Contingencies" team members.

References

[1] Hartzband P, Groopman J. Off the Record — Avoiding the Pitfalls of Going Electronic. N Engl J Med [Internet]. Massachusetts Medical Society; 2008 Apr 17;358(16):1656–8. Available from: http://dx.doi.org/10.1056/NEJMp0802221

[2] Campbell EM, Sittig DF, Guappone KP, Dykstra RH, Ash JS. Overdependence on technology: an unintended adverse consequence of computerized provider order entry. AMIA Annu Symp Proc [Internet]. 2007 Jan [cited 2014 Jul 24];94–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2710605&tool=pmcentrez&rendertype=abstract.

[3] Liker JK, Haddad CJ, Karlin J. Perspectives on technology and work organization. Annu Rev Sociol. JSTOR; 1999;575–96.

[4] Gérvas J. Longitudinalidad y continuidad de la atención al paciente y a la familia: ayuda tecnológica a un problema omplejo sin soluciones sencillas [Internet]. Madrid: Equipo CESCA,; 2008; 1. Available from: http://www.fcs.es/docs/jornadas/politicasanitaria/documento_debate_Juan_Gervas.pdf.

[5] Galvez Ibañez M. Continuidad Asistencial. Análisis conceptual de los actores y amenazas. Propuestas y alternativas. Medicina Familiar (And) Vol 4 N° 1; 2003; p. 58.

[6] Swanson M, Wohl A, Pope L, Grance T, Hash J, Thomas R. Contingency planning guide for information technology systems: Recommendations of the National Institute of Standards and Technology. DTIC Document; 2002.

[7] ISO 22301:2012 - Societal security - Business continuity management systems - Requirements [Internet]. 2012 [cited 2014 Jul 24]. Available from: http://www.iso.org/iso/catalogue_detail?csnumber=50038.

[8] The IT Governance Institute [Internet]. IL; USA: IT Assurance Guide: Using COBIT; 2007; 62. Available from: http://www.itgi.org/

[9] Fernández Romero A. Manual del Consultor de Dirección. Madrid: Ediciones Díaz de Santos; 2008. p. 177.

[10] Nelson NC. Downtime procedures for a clinical information system: a critical issue. J Crit Care [Internet]. 2007 Mar [cited 2014 Jul 24];22(1):45–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17371746.

[11] Menon S, Singh H, Meyer AN, Belmont E, Sittig DF. Electronic health record-related safety concerns: a cross-sectional survey. J. A. S. for Health Risk Manag. 2014;34(1):15.

[12]Sittig DF, Gonzalez D, Singh H. Contingency planning for electronic health record-based care continuity: A survey of recommended practices. Int J Med Inform. 2014 Nov; 83(11):803. doi: 10.1016/j.ijmedinf.2014.07.007.

[13]Lei J; Guana P; Gaoa K; Lua X; Chenc Y; Li Y, Mengd Q; Zhangb J, Sittig D; Zhenge K. Characteristics of health IT outage and suggested risk management strategies: An analysis of historical incident reports in China. Int J Med Inform. 8 3, 2014; 122–130.

[14]Fahrenholz C, Smith LJ, Tucker K, Warner D. Plan B: A Practical Approach to Downtime Planning in Medical Practices. AHIMA. 2009;80(11):35.

[15]James, E.M.; Hangiandreou N.; Bender C. Electronic imaging system implementation at Mayo Clinic Rochester: downtime procedures and communication plans. J Digit Imagin. 2000;7.

Address for correspondence:

Marcela T. Fernández, M.Sc.

Juan D. Perón 4190 (C1199ABD), Edificio DIS,

Ciudad Autónoma de Buenos Aires, Argentina

teresa.fernandez@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-477

Flexibility First, Then Standardize: A Strategy for Growing Inter-Departmental Systems

Arnvør á Torkilsheyggi[a]

[a] *Computer Science and Informatics, Roskilde University, Roskilde, Denmark*

Abstract

Any attempt to use IT to standardize work practices faces the challenge of finding a balance between standardization and flexibility. In implementing electronic whiteboards with the goal of standardizing inter-departmental practices, a hospital in Denmark chose to follow the strategy of "flexibility first, then standardization." To improve the local grounding of the system, they first focused on flexibility by configuring the whiteboards to support intra-departmental practices. Subsequently, they focused on standardization by using the whiteboards to negotiate standardization of inter-departmental practices. This paper investigates the chosen strategy and finds: that super users on many wards managed to configure the whiteboard to support intra-departmental practices; that initiatives to standardize inter-departmental practices improved coordination of certain processes; and that the chosen strategy posed a challenge for finding the right time and manner to shift the balance from flexibility to standardization.

Keywords:

Standardization; Electronic Whiteboards; Organizational Implementation.

Introduction

The design and implementation of IT systems in hospitals has received increased attention as a way to improve healthcare through standardization [1, 10]. The standardization of health data facilitates comparison between departments and hospitals which subsequently allows decision-makers to make informed decisions about best practices [11], while standardization of clinical practice has shown to improve quality and reduce patient harm [9]. The reach of standards can vary; some have global reach and have already been crafted (e.g. International Classification of Diseases), while others are national, regional or even more local (e.g. a department developing a list of diagnoses for their own use). Clinicians will often perceive standardization as an unwelcome constraint on their practice, and any attempt to standardize practices through the design and implementation of IT systems is faced with the challenge of balancing the need for standardization against the clinicians' need for a technology that can support flexibility of use [2, 5]. Some have proposed to find this balance by first focusing on standardization and then configuring the standards to the local context, or in other words: first standardize, then localize [2].

This paper reports a case where the reverse approach was chosen. The case consists of the development of an electronic whiteboard system for hospitals in Region Zealand, Denmark. The system was designed in collaboration between a Norwegian vendor, Region Zealand, and Roskilde University. In 2009, the system was implemented at the four Emergency Departments in the region with the purpose of supporting their maintaining an overview of the patients at the department. The system is web-based and the basic layout of the whiteboards can be likened to a spread-sheet, with a row for each patient and columns displaying selected information about the patient such as triage level, room, chief complaint, responsible nurse, responsible physician, and current treatment activity (see Figure 1).

In the beginning of 2013 the whiteboard system was introduced at all departments in one of the hospitals in the region. The purpose of this implementation was to support and improve inter-departmental practices through standardization.

Figure 1 – The electronic whiteboard system.

Because the whiteboard system is highly customizable, the whiteboards could be configured to fit and support the practices in the wards. In order to circumvent resistance of a top-down standardization attempt, it was decided to postpone standardization attempts and to focus on securing the local grounding of the system in the wards. Selected super users were appointed to be responsible for the implementation on their respective wards, which included configuring the system to support their intra-departmental practices. These super users were shown pre-configured templates (e.g. Figure 1) and were then trained to re-configure the templates; by adding or removing columns, deciding how data should be inserted and displayed and so forth. In addition, they were trained in how to configure views or filters that the other users could activate by clicking on a button; for example if the users wished to only see patients with a certain problem, or only patients in a certain part of the ward.

The approach in the beginning of the implementation primarily focused on flexibility and configuring the whiteboards to local contexts. When the system had been configured to support intra-departmental practices and had been in use for a while, the management facilitated activities where clinicians from different departments were called together to use the whiteboards as a starting point to negotiate standardization of inter-departmental practices. This paper investigates to what extent the chosen approach of "flexibility first, then standardize," which in the managers' perspective has been a viable approach for finding a balance between standardization and flexibility in the implementation process.

Related Work

This section summarizes previous work in the following three areas: 1) the benefits and potential of standardization within healthcare; 2) the importance of developing technologies that support flexibility of use; and 3) the challenge of finding a balance between standardization and flexibility.

Several studies have pointed to the potential benefits of standardization. Some have stressed the importance of standardizing health data in order to enable comparisons and thus to allow decision-makers to make informed decisions about best practices [11]. Others have found that the standardization of practice through protocols and checklists has shown to improve quality and reduce patient harm [9]. It has also been argued that the development and implementation of IT holds the potential to improve quality and increase patient-safety through the standardization of otherwise error-prone work processes [1].

When designing and implementing technology, researchers have increasingly acknowledged that it is impossible to design a perfect system prior to actual use [2]. This acknowledgement has motivated the development of approaches that allow for local customization and reconfiguration to be performed after the implementation of the technology, when the users have had the opportunity to use the technologies in their real work [3, 6, 7]. It follows that in order to support customization and reconfiguration, the implemented technology must be configurable and flexible; in other words, it must allow for continued design-in-use [7]. Flexibility in this sense refers to developing technology that allows for further changes and reconfigurations as well as flexibility in the patterns of use [5]. However, most design-in-use approaches do not explicitly address the challenges that arise when the agenda of allowing flexibility clashes with the agenda of standardizing practices across departments, specializations and staff-groups.

Although IT can potentially support standardization, clinicians may perceive the IT as a constraint on the way they usually go about their work on various wards. This is a well-known issue within the research area of information infrastructures [2]. In order for an inter-departmental IT system – like the one described here – to be adopted by the clinicians, it has to be useful in the local context [4]. The agenda of standardization inevitably is faced with the challenge of finding a balance between standardization and flexibility [2]. Finding this balance is, however, no easy task. In their study of the customization process of an electronic triage and tracking system that was configured to be used in eight Canadian emergency departments Bjørn et al. [2] realized that it was impossible to reach a shared (standardized) solution without constraining crucial work in the departments. They therefore proposed to find the balance between standardization and flexibility by first having actors from different departments focusing on how much can be standardized without constraining local flexibility and then focus on configuring the standardized template to the local context. In other words, they proposed the approach: first standardize, then localize [2]. The approach reported in this paper suggests the opposite: focusing on flexibility first and then standardization. In the following discussion, the chosen approach will be discussed and evaluated in relation to the approach proposed by Bjørn et al.

Method

The study was conducted at a medium sized hospital in Region Zealand, Denmark, where the author has followed the implementation and use of the whiteboard system for two years. The study was approved by the hospital management. The findings are primarily based on semi-structured interviews with managers that on different levels have been responsible for managing the implementation process. Table 1 provides an overview of the interviewed managers and their responsibilities. Local Coordinator 1 (LC1) was a nurse with special qualifications who in September 2012 was asked to be a system administrator and to support the implementation process. She held this position until December 2013. At the time of our study, LC2 was the local coordinator.

Table 1 – List of Interviews

Position	Responsibilities
Deputy Director of the Hospital	Head of the local steering committee
Regional Project Manager	Managing the regional project
Local Coordinator 1	Managing the local implementation process
Local Coordinator 2	Managing the local implementation process
Chief Physician	Managerial responsibilities within the surgical department

The interviews were semi-structured. An interview guide was sent to the interviewees prior to the interviews. The guide mainly revolved around the topics: what have the challenges been in regards to managing the implementation process and how have they been dealt with; what initiatives have been taken in order to achieve standardization; how have you dealt with the challenge of balancing standardization and flexibility.

All five managers agreed to have the interviews audio-recorded. The interviews were subsequently transcribed and

analyzed according to a grounded theory approach. Each interview was coded using an open-coding process. The codes were registered in a spreadsheet along with a short description as well as an indication of which interview they originated from. The initial codes were then arranged and re-arranged through an iterative process until the three overall categories emerged: 1) Flexibility and the local: The facilitation of a user-driven process focusing on local grounding to support intra-departmental coordination; 2) Standardization and the global: The need for managerial support when standardizing inter-departmental coordination; and 3) The challenges of balancing these two concerns.

Results

Flexibility to Achieve Local Grounding

Even though the purpose of the implementation was to support and standardize inter-departmental practices, it was decided to focus first on achieving the local grounding of the system through a user-driven process. The message that the management wanted to send to the users was: *"We have been saying: we know that you are doing a great job. We know that you have the knowledge, competences and the abilities to make this even better."* (project manager). The initial focus therefore concentrated on supporting the wards in configuring the system to support intra-departmental practices.

A nurse with special qualifications from the surgical department was appointed as local coordinator. She was trained to be a system administrator and was asked to support the implementation process and to train selected users in configuring and using the system. During the fall of 2012, hospital management asked the hospital department managers to appoint 2-3 clinicians from each ward. These clinicians were given the responsibility of being super users. The super users' responsibility was to support the implementation process in their respective wards by configuring the whiteboard and teaching their colleagues how to use it. The process of training the super users to take on this responsibility started with workshops, where they were trained in configuring the whiteboard. After these introductory workshops, super users were further encouraged to attend additional workshops, where they could come and receive support in configuring their whiteboards. These workshops were, however, optional and only the super users from some of the wards chose to accept the offer. When asked why these workshops were optional, the local coordinator replied: "*We wanted to say that we really do not have an agenda. Because you should bring your agenda to us! We are just here to help you, if you need help with changing the columns, making filters etc."* (LC1). Because the management chose to focus on flexibility first, the super users were given free reins as to how they chose to configure the whiteboards. According to the local coordinator, this strategy was chosen because it was expected to increase the chances of a successful implementation if the super users (and the other clinicians in the ward) felt like they could influence the appearance of the whiteboards. However, because the system was intended to support inter-departmental communication, it was important to secure a basic level of standardization. There were, therefore, some columns that the super users were required to include in their configuration of the whiteboard. These columns were referred to as "global columns" and displayed information like social security number, name, problem and plan. For some of these global columns, it was up to the super users to decide what and how data should be registered in them. When deciding how the problem-column should be used, some super users created a list of problems that were typical for their ward and the end-users were then required to choose one of these predefined problems when entering data into the column. On other wards, end-users were allowed to write free-text in the problem-column.

The super users were encouraged to configure the whiteboard system to support their intra-departmental practices by focusing on their daily routines. The interviewed chief physician explained how he and the other super users from his ward had solved this task: *"Nurses, secretaries and I […] defined what we wanted to see in the morning. Those 20-25 patients that we were going to treat. We wanted to get an overview. You had to look at your workday and say: how do we configure this to our work while simultaneously realizing that you may need to change your organization."* (Chief physician). This quote shows that the super users were not oblivious to the fact that even though they focused on configuring the whiteboards to their existing practices, they also realized that the whiteboards would potentially change the way they went about their work. The chief physician explained that he saw this configuration process as a fantastic opportunity, because it was the first time that he had participated in an implementation process where he had an opportunity to say what he wanted the system to look like. According to the local coordinator, many wards succeeded in configuring the whiteboards to support their intra-departmental practices.

Increased Standardization

It did, however, become increasingly clear to the clinicians that there was a need for increased standardization, if the system was to be used to support inter-departmental practices. Even though the super users had been required to include the global columns in order to enable inter-departmental communication, they could – if they felt that one of these columns was less important on their ward – choose to configure their whiteboards in such a way that these columns were not easily seen (and therefore typically ignored) by the clinicians on their ward. Additionally, most wards made frequent use of the predefined views that the super users had configured and these views only displayed selected columns. Because the whiteboards differed from ward to ward, clinicians in one ward did not know what clinicians in others wards actually saw: "*Everybody was like: I want it to be like this and I want the columns to be there and the views to search for that and that. And yes, the message was: you can get what you want. But when we wanted to communicate then we discovered that we couldn't. The information that I put into the system did not reach the intended receiver."* (chief physician). The chief physician pointed out that because of the lack of standardization with regards to columns and views, the super users realized that if they wanted to ensure that the information on the whiteboard reached the intended receiver, they had to cooperate with super users in other wards: "*If anyone wanted to inform another ward or group about something, then you had to say [to the super user in the other ward]: you have to modify your filter to include information from this field."* (chief physician). Thus, even though information was entered into the system with the intent of sharing it with another ward, the lack of standardization in terms of the appearance of the whiteboard meant that the system in practice could not be used to support inter-departmental communication.

As it became increasingly clear to the clinicians that there was a need to increase standardization, the management also became increasingly aware that it was time to constrain the flex-

ibility that had been given to the wards. Therefore, super users and management across departments gathered to negotiate and agree on rules about how to configure and use the whiteboards to support inter-departmental communication.

It was decided to take a more top-down approach in order to initiate the standardization of selected inter-departmental work processes: "*We said: well, let us try to focus on certain work processes and then get some common ground regarding how we want it to be.*" (LC1). The first example of the standardization of such an inter-departmental work process concerned the boarding of patients for surgery. The local coordinator convened managers and super users from the relevant departments to discuss how the boarding should be supported by the whiteboard: "*And they really got up to a discussion (...) and I thought, how can I close this discussion! (...) But in the end of the first meeting, they were about 80% in agreement.*" (LC1). After a few more meetings, the involved clinicians agreed on a standard checklist that entailed 7 steps that were required in order to secure a safe transfer for patients that needed surgery. When all 7 items on the checklist had been checked, the patient could be transferred from the ward to the OR. The local coordinator configured the whiteboard to include this checklist to support the agreed inter-departmental work process. The standardization of the boarding of patients for surgery thus required super users and the management from several relevant departments to agree on a standard. When this standard had been agreed upon however, the work of implementing the standard still remained: the clinicians in the departments had to be trained in adhering to the standard and using the whiteboard in a structured way. For a few weeks, the clinicians could use both the old way of boarding patients (fax and phone) and the new way through the whiteboard. After a trial period, it was decided by the hospital management that all patients had to be boarded through the whiteboard. The process of using the whiteboard as a starting point to standardize inter-departmental work processes has subsequently also been used in other areas (for example for ordering physiotherapists). Presently, the hospital is pilot testing a new checklist in the whiteboards to be used in the referral of patients from the ED to other departments. The new checklist has also been designed based on a standardized practice that clinicians from relevant departments have agreed upon. The region now plans to implement these standards in other hospitals in the region, where the whiteboards are presently being implemented.

Another important issue in regard to standardization has to do with the standardization of data and how the whiteboards can be used to generate statistics about the workflow in departments and hospitals. Presently, the region is working to generate regional standards with regards to the content that is registered within the columns. This includes the generation of lists of classifications that the clinicians can choose from rather than typing free-text. According to the present local coordinator, there has been a wish to work on these classifications for a long time, but the whiteboards have been a catalyst for focusing on and negotiating much needed standardization across departments and hospitals in the region.

Balancing Flexibility and Standardization

When the whiteboards were implemented, the hospital management chose to pursue a bottom-up approach to improve the chances of local grounding. In this first phase, the focus was on flexibility, where the super users seemingly were given free rein to configure the system to support intra-departmental practices. When local grounding had been achieved, the management decided to proceed with the actual goal of standardizing inter-departmental practices. Several of the interviewed managers reported how it had been a challenge to make the transition from flexibility to standardization: *"When do you go from saying: use it, if it makes sense for you – to saying: Use it! When do you make that switch?"* (project manager). The deputy director felt that the transition was dependent upon whether the system was in use in the wards: *"When the basic use of the system is widely grounded in our organization (...) then we need to cut back on the free reins and work to control how we use the whiteboards."* (deputy director). Thus, the management should not focus on standardization until the system was being used to a sufficient degree in the wards.

The interviewed managers stressed that the standardization of inter-departmental practices had been highly dependent upon the engagement of the top management in the hospital. The deputy director stated that he too was aware of the importance of the top management's engagement and therefore recently made an official announcement that all departments are required to use the whiteboards to a certain degree and for selected practices. According to the initial local coordinator, the work to standardize inter-departmental practices prompted a shift in the managerial approach: "*It resulted in the use of some coercion. When we put these work processes into the system, we removed the old way of doing it. It was like saying: It is no longer a question of whether you want to or not. This is now the way to do it!*" (LC1). In order to agree on inter-departmental standards and to get the clinicians to adhere to them, the hospital management and the local coordinator chose to use some coercion and hence to constrain the flexibility that had been central in the first phase of the process.

The shift from flexibility to standardization also brought forth a shift in challenges. The deputy director explained that when the implementation was driven by super users, the main challenge was to reach decisions about inter-departmental practices. He did however also see the approach as a promising investment: "*It has some clear advantages in that you get to develop things that are clinically relevant.*" (deputy director). With a more management-driven process, it is easier to make decisions about inter-departmental issues but it can be a challenge to foster good (clinically relevant) ideas: *"How do we make sure that we foster good ideas? How do we make sure that we realize what the best practice actually is?"* (deputy director). He therefore saw the importance of continued close cooperation with super users on every ward and setting up the proper forums to secure a bottom-up flow of knowledge and good ideas from the clinicians (in terms of how the whiteboards could be used to develop and standardize intra- and inter-departmental practices).

Discussion

As shown in the preceding section, the chosen strategy for the implementation of the whiteboards, in terms of finding a balance between standardization and flexibility, was one of "flexibility first, then standardize." The following discussion will focus on whether this approach was a viable alternative to Bjørn et al. [2] approach of first standardize, then localize.

In the first phase, the management chose to focus on flexibility by facilitating a user-driven approach and refraining from setting too many criteria for the super users. On many wards, this meant that enthusiastic super users welcomed the configurability of the system as an opportunity to influence both the appearance and the use of the system in their daily work. The decision to achieve local grounding before standardization

meant that the whiteboards were perceived as useful by the clinicians on wards, where super users had successfully configured the whiteboards. When the implementation entered the second phase, standardization, clinicians had already adopted the whiteboards in their local context and were becoming increasingly aware of the need for negotiating standards across wards. Even though this standardization process meant that some super users felt that they had to roll-back some of their changes, they acknowledged the standardization as a necessity if the whiteboards were to support inter-departmental coordination.

Due to the management in the first phase (flexibility) choosing not to set (or reveal) their agenda or boundary conditions for the change process, the super users in some wards never managed to incorporate the whiteboards into local practices; they simply did not know what was expected of them. Several of the managers acknowledged that the users had been given too much flexibility and too little direction. Their explanation was that the hospital management wanted the process to be user-driven. This however points to a somewhat misunderstood notion of what is required to manage a user-driven process. Within the tradition of Participatory Design, which is one of the central user-driven disciplines, it has been highlighted as to how important it is that the participants in the design have "close links to project management" [8]. According to Simon [12], an open-ended design process requires more of the users than a process with clear directives and she therefore argues for helping users in user-driven processes by "scaffolding" their participation. The management of user-driven processes thus requires the provision of supportive resources, tasks, and guidance upon which the super users can build their confidence and abilities.

When the management took more control in order to address standardization, they stated that there was a shift, where the process could no longer be driven by users; rather, the users had to be told what to do and how to use the whiteboards. In the light of the aforementioned scaffolding perspective, one might argue that both flexibility and standardization rely on the management scaffolding (super) users' participation in terms of how to use the whiteboards to improve and develop intra- as well as inter-departmental practices. When the deputy director stressed the importance of setting up forums to secure a continued bottom-up flow of knowledge and good ideas from clinicians, he was basically talking about how to scaffold a user-driven change process.

Conclusion

When implementing IT in an attempt to standardize inter-departmental practices, this study suggests that it can be a viable strategy to focus on flexibility first, then standardization. The first phase, flexibility, was succesful in terms of configuring the whiteboards to secure local grounding in most wards. In hindsight, however, several managers stated that the users had been given too much flexibility and too little direction. In the second phase, standardization, it was an advantage that the whiteboards had been widely adopted and that users experienced the need for standardization. The strategy also poses challenges in terms of managing the user-driven process, scaffolding the clinicians' participation, and in finding the right time and manner to transition from flexibility to standardizaiton.

Acknowledgments

This study was co-funded by Region Zealand as part of the Clinical Communication project. Special thanks are due to the interviewed managers.

References

[1] Bates, D.W. The quality case for information technology in healthcare. *BMC Medical Informatics and Decision Making*. 2002: 2(7): 1–9.

[2] Bjørn, P., Burgoyne, S., Cropton, V., MacDonald, T., Pickering, B, and Munro, S. Boundary factors and contextual contingencies: Configuring electronic templates for healthcare professionals. *European Journal of Information Systems*. 2009: 18(5): 428–441.

[3] Botero, A., Kommonen, K-H., and Marttila, S. Expanding design space: Design-in-use activities and strategies. In: *Proc. of DRS* 2010. Montreal, Canada.

[4] Ellingsen, G. Coordinating work in hospitals through a global tool. *Scandinavian Journal of Information Systems*. 2003: 15: 39–54.

[5] Hanseth, O., Monteiro, E., and Hatling, M. Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology & Human Values*. 1996: 21(4): 407–426.

[6] Hartswood, M., Procter, R., Slack, R., Voss, A., Buscher, M., Rouncefield, M., and Rouchy, P. Co-realisation: Towards a principled synthesis of ethnomethodology and participatory design. *Scandinavian Journal of Information Systems*. 2002: 14(2): 9–30.

[7] Henderson, A. and Kyng, M. There's no place like home: Continuing design in use. In: Greenbaum, J. and Kyng, M., eds, *Design at Work: Cooperative Design of Computer Systems*. Erlbaum. 1991: 219–240.

[8] Kensing, F., Simonsen, J., and Bødker, K. MUST: A method for participatory design. *Human–Computer Interaction*. 1998; 13(2): 167–198.

[9] Kirkpatrick, D.H. and Burkman, R.T. Does standardization of care through clinical guidelines improve outcomes and reduce medical liability? *Obstetrics and gynecology*. 2010: 116(5): 1022–6.

[10] Raghupathi, W. Health care information systems. *Communications of the ACM*. 1997: 40(8): 80–82.

[11] Richesson, R.L., and Krischer, J. Data standards in clinical research: Gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association*. 2007: 14(6): 687–696.

[12] Simon, N. 2010. *The Participatory Museum*. Museum 2.0.

Address for correspondence

Arnvør á Torkilsheyggi, Computer Science and Informatics, Roskilde University, Bldg. 43.2, DK-4000, Roskilde, Denmark, email: arnvoer@ruc.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-482

Human Factors Predicting Failure and Success in Hospital Information System Implementations in Sub-Saharan Africa

Frank Verbeke[a,b], Gustave Karara[a], Marc Nyssen[a,b]

[a] *Department of Biostatistics and Medical Informatics, Vrije Universiteit Brussel, Brussels, Belgium*
[b] *Regional e-Health Center of Excellence, University of Rwanda, Kigali, Rwanda*

Abstract

From 2007 through 2014, the authors participated in the implementation of open source hospital information systems (HIS) in 19 hospitals in Rwanda, Burundi, DR Congo, Congo-Brazzaville, Gabon, and Mali. Most of these implementations were successful, but some failed. At the end of a seven-year implementation effort, a number of risk factors, facilitators, and pragmatic approaches related to the deployment of HIS in Sub-Saharan health facilities have been identified. Many of the problems encountered during the HIS implementation process were not related to technical issues but human, cultural, and environmental factors. This study retrospectively evaluates the predictive value of 14 project failure factors and 15 success factors in HIS implementation in the Sub-Saharan region. Nine of the failure factors were strongly correlated with project failure, three were moderately correlated, and one weakly correlated. Regression analysis also confirms that eight factors were strongly correlated with project success, four moderately correlated, and two weakly correlated. The study results may help estimate the expedience of future HIS projects.

Keywords:

Hospital information systems; Biomedical technology assessment; Computer systems evaluation

Introduction

HIS implementation has gained momentum in Sub-Saharan Africa in the past decade. In the period from 2007 to 2014, the authors participated in the implementation of the open source OpenClinic GA HIS [1] in 19 hospitals in Rwanda, Burundi, DR Congo, Congo-Brazzaville, Gabon, and Mali. Many of these implementations have been successful, but some of them turned into failures. Introducing HIS proved to be a complex process. There was no single standard pathway to successful implementation. Sometimes what failed in one health facility seemed to work very well in another one. In the beginning, many solutions had to be discovered by trial and error. Nevertheless, at the end of a seven-year implementation effort a number of risk factors, facilitators, and pragmatic approaches related to the deployment of HIS in Sub-Saharan health facilities have been identified. These elements have been categorized in the following classes:

- Infrastructure
- Patient administration
- Financial information management
- Reason for encounter and diagnostic coding
- Medical record management
- Lab information management
- Medical imaging
- Reporting and statistics
- Systems integration
- Project management issues and human factors

Many of our findings were in line with earlier publications that shed light on isolated aspects of health information system issues [2,3,4,5] and environmental country-specific problems [2,6,7]. Many of the issues that were encountered during the OpenClinic GA implementation process were not technical, but could be brought down to human, cultural, and environmental factors. Although such issues were frequently irrational and hard to solve, they seemed to be fairly predictive in determining the success or the failure of a project. In order to formally evaluate the predictive value of these human factors in our HIS implementations in the Sub-Saharan region, a retrospective, descriptive, and semi-quantitative study has been set up.

Materials and Methods

In the course of the 19 HIS implementation projects, implementation logbooks have systematically been kept that document a wide range of problems that arose, solutions, and workarounds, as well as the results that were obtained from implementing the solutions. Analysis of the logbooks—which included feedback from the local project managers in the 19 health facilities—resulted in the identification of the aforementioned 10 different classes of issues. The *Project Management Issues And Human Factors* class consisted of a list of 14 factors that could potentially be related to project failure and 15 factors that were potential candidates for predicting sustainable implementation success.

A global project implementation success score ranging from 0 (complete failure) to 5 (complete success) has been awarded to each of the 19 health facilities in our research. This score was derived from the level of agreement expressed by the hospital management staff and an external evaluator with a series of six statements ranging from 0 (completely disagree) to 5 (completely agree):

1. All goals which have been set out in the project scope have been achieved.
2. All project results have been delivered in time.
3. All intended users have been trained and are using the system.

4. The health information system implementation contributed to the improvement of health facility productivity.
5. The health information system has contributed to quality improvement of health care services.
6. The health facility is able to self-support post-project operational costs related to the health information system.

For every health facility, a score was then allocated representing the relevance of each of the 14 failure and 15 success factors. Scores ranged from 0 (irrelevant) to 5 (of highest importance) and the scoring of each factor was the result of a consensus between local health facility management and an external implementation evaluator. An average *risk level* score was then calculated for each health facility based on the average failure factor scores that were previously allocated. Similarly, the average score for the success factors was called the *opportunity level*. Finally, correlations were calculated between individual factors and project success scores.

Results

Implementation Failure Factors

Several authors have published extensive lists of potential pitfalls and failing factors for HIS deployment in developing countries. In 2002, Heeks [8] drew a rather pessimistic view of the health informatics landscape in the developing world mainly based on a perceived mismatch between information systems design and local user actuality. In our HIS implementations, we specifically tried to address issues related to such implementation gaps. Failure factors identified in our research substantially differed from Heeks' and other authors' [4,5] observations. What follows is a tabular summary of a number of factors that could be predictive for eventual project failure (Table 1). Predictive values show the correlation *r* with global project success and the *p*-value for the F-test on the regression with a confidence level of 95%. Significant correlations below 0.60 are labeled as having *low* predictive value, between 0.60 and 0.75 *moderate* predictive value, and above 0.75 *high* predictive value.

Table 1 – Project failure factors evaluation.

Failure factor	Description	Predictive value
1. Unclear goals	When no clear definitions of an intended outcome exist, measurement of success becomes extremely difficult. False and unrealistic user expectations are very likely to generate a perception of project failure.	r = 0.70 p<0.001 moderate
2. Absence of a project champion	Project champions are extremely important for aligning user behavior to set out project scope. Absence of a project champion often leads to disinterest and lack of user motivation.	r = 0.70 p<0.001 moderate
3. Resignation to poor health service quality	Hospital management resigning to the poor health status of patients and accepting inadequate health care practice are bound to fail in implementing information systems. Implementations can only be justified by health services improvement.	r = 0.81 p<0.0001 high
4. Psychological factors	Initial, often irrational user resistance and physician skepticism are predictive for slower implementation progress.	r = 0.78 p<0.0001 high
5. Organizational culture and silo mentality	Health facilities composed of individual relatively autonomous departments with minimal perception of system-wide collective goals carry a risk of fractional implementation. Lack of trust can also play a pernicious role.	r = 0.76 p<0.001 high
6. Resistance to change and power shifts	Documented resistance to change based on fear of losing a job or an advantageous position are often present in insecure working environments. If no measures are being taken to address such issues, partial or total project failure is more likely to occur.	r = 0.74 p<0.001 moderate
7. Time	Unrealistic implementation timeframes are bound to fail. Another time factor relates to the fact that information and communication technology (ICT) tools should never lengthen the time needed for performing routine operations. Systems that reduce users' productivity will rarely survive.	r = 0.90 p<0.0001 high
8. Poor technical assistance	In several sites, perceived technical assistance quality was directly correlated with overall users' system appreciation. Effective technical support is key to information systems being seen by users as dependable working instruments.	r = 0.80 p<0.0001 high
9. Insufficiently skilled staff	Health facility staff with extremely low ICT skills are somehow predictive for failure as they add an	r = 0.59 p<0.01 low

	additional burden to training and user assistance, and increase funding requirements.	
10. Insufficient training	Inadequately trained users provide inadequate results. Although such logic does not apply to ICT-based systems exclusively, often ICT-training inadequacy is being translated into excessive system complexity. Additionally, continued training should address staff turnover.	r = 0.77 p<0.0001 high
11. Discontinued follow-up	Users being left alone after initial start-up training creates a sense of disinterest. Result-oriented implementations must address the need for individual follow-up.	r = 0.90 p<0.0001 high
12. Poor mapping on prevailing practices	ICT tools which are poorly mapped onto existing routines and practices will be abandoned in the end. If users fail to find personal interest in developed instruments, they will not use them.	r = 0.27 p = 0.26 NA
13. Perceived complexity	Modules that contain a high level of perceived complexity will frighten many users and induce unnecessary resistance to change. ICT skills being scarce in Sub-Saharan Africa, complicated systems must be avoided.	r = 0.77 p<0.001 high
14. Low user satisfaction	Low user satisfaction in the course of project implementation must always be considered an alarming element. In the case of clear user dissatisfaction, ongoing implementation should be reconsidered.	r = 0.97 p<0.0001 high
Global risk level	Average score for all of the above failure factors.	r = 0.95 p<0.0001 high

Based on our study set, 9 of the 14 factors were strongly correlated with project failure, 3 factors were moderately correlated, and 1 factor had low predictive value. The evaluation of factor 12 (poor mapping on prevailing practices) produced no valuable output because too few health facilities faced this kind of issue. The calculated global risk level score demonstrated a strong, statistically-significant correlation with project success (Figure 1).

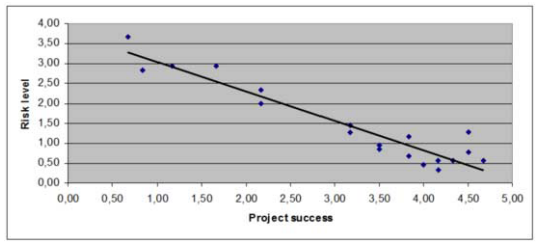


Figure 1 – Risk level and project success regression.

Implementation Success Factors

Similar to the identification of failure risks, a list of elements that increase the likelihood of successful project implementation have also been isolated. A number of such factors has already been described by Hendricks [4] based on a review of 31 reports on electronic health records implementations in developing and developed countries. In our study results we found that many of the factors in Hendricks' report did not fit the Sub-Saharan healthcare reality. What follows is a tabular summary of a shorter list of success factors that applied to our own research experience (Table 2).

Table 2 – Project success factors evaluation.

Success factor	Description	Predictive value
1. Broad staff enrollment	When the majority of users at all levels of the health facility have the feeling of participating in the implementation effort, HIS can develop beyond the stage of being a management whim.	r = 0.68 p<0.01 moderate
2. Clear communication	Clear, broad communication of the goals and the potential impact of the new technology on work organization helps to cope with resistance to change and distrustful mindsets.	r = 0.85 p<0.0001 high
3. Realistic timing	Timing should be realistic enough to address reasonable stakeholders' expectations. A well-planned implementation with a feasible schedule assures better synchronization between different projects' tasks.	r = 0.80 p<0.0001 high
4. Progressive change management	Big bang implementations where massive business process change is pushed through complex organizations, such as health facilities, is a dangerous practice. Gradual and progressive introduction of new modules often appears more successful.	r = 0.70 p<0.001 moderate

5. Incentives	Motivation of HIS users in the form of an official recognition, a small reward, or even a simple pat on the back have been found mainly in successful projects.	r = 0.43 p<0.07 Not significant
6. Business process reengineering ability	The ability of a health organization to cope with necessary transformations of business processes that get in the way of information systems implementation is important for achieving the best productivity improvement.	r = 0.81 p<0.0001 high
7. Stakeholder consensus	Optimally, HIS implementation must be in line with the goals of all stakeholders. Reorientation of project scope must therefore always be based on stakeholder consensus.	r = 0.75 p<0.001 high
8. Holistic approach	Systems that efficiently integrate information from different departments better contribute to global business targets, such as generic patient health status improvement and hospital productivity.	r = 0.82 p<0.0001 high
9. *Quick wins* implementation	*Quick wins* can constitute an essential step in bridging the start-up gap between systems' potential and users' mistrust. They can greatly motivate users for continuing to invest in a project.	r = 0.91 p<0.0001 high
10. Sufficient and continued training	Users' needs for training must always be considered seriously. Users who feel comfortable with the system's operating procedures show a more positive attitude towards the project, resulting in better outputs.	r = 0.68 p<0.01 moderate
11. Clinician and ICT staff intermediaries	The presence of health IT experts can resolve many communication errors between IT technicians and clinical staff pro-actively. This can reduce frustration and inadequate allocation of development time.	r = 0.54 p = 0.02 low
12. Consideration of prevailing practice	HIS should never dictate users' business processes, but rather adapt to the actuality of the activity they support. Doing so can transform these systems into highly valued instruments that facilitate work.	r = 0.56 p = 0.01 low
13. Adequate technical assistance	Many users don't mind facing technical problems once in a while, as long as these can be quickly and effectively addressed by technical support staff. Unresponsive support staff can compromise project survival.	r = 0.85 p<0.0001 high
14. High measured user satisfaction	Obviously, a high level of user satisfaction being measured in the course of project implementation is a strong indicator of important user concerns being appropriately addressed.	r = 0.94 p<0.0001 high
15. Perceived user-friendliness	Simplicity and suitability of user interfaces for performing frequent tasks are generally perceived as user-friendly, which improves user acceptance and reduces initial resistance to change.	r = 0.69 p<0.01 moderate
Global opportunity level	Average score for all of the above success factors.	r = 0.91 p<0.0001 high

Regression analysis confirms 8 factors are strongly correlated with project success, 4 factors have a moderate correlation, and 2 factors have a weak correlation. The average opportunity level also strongly and significantly correlates with project success (Figure 2). For factor 5 (incentives), no statistically significant relationship to eventual project success can be demonstrated.

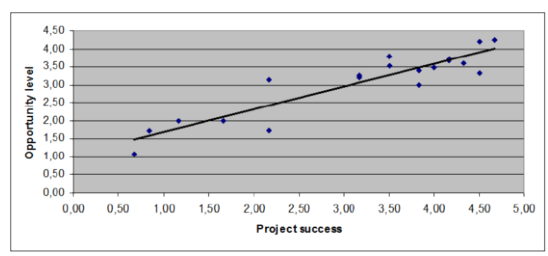


Figure 2 – Opportunity level and project success regression.

Conclusion

The developed failure risk assessment scheme confirmed, to a large extent, the relevance of the failure risk factors we extracted from many interviews with several hundred health facility users. As expected, the user-oriented nature of the OpenClinic GA HIS led to results that were substantially different from data that had been previously published by other authors [4,8]. This mainly resulted in a low predictive value of risk and success factor 12 (mapping on prevailing practices) in our study sample.

Although every health facility constitutes a specific case and no absolute information can be derived from the listed risk and success factors, they may offer some grip for estimating the potential expedience of future HIS projects that are being considered.

References

[1] Verbeke F, OpenClinic GA, http://sourceforge.net/projects/open-clinic
[2] Benson AC, Adoption of Hospital Information Systems in Nigeria. Journal of Global Health Care Systems. 2011;1(3):1–38.
[3] Malik MA, Khan HR. Understanding the Implementation of an Electronic Hospital Information System in a Developing Country: A Case Study from Pakistan. Wellington, New Zealand: Conferences in Research and Practice in Information Technology (CRPIT); 2009.
[4] Hendricks F, Factors to Consider when Planning an Electronic Health Record (EHR) System Implementation: Global Lessons for South Africa. University of Oregon; 2012.
[5] Bukachi F, and Pakenham-Walsh N. Information Technology for Health in Developing Countries. Chest. 2007 Nov 1;132(5):1624–30.
[6] Abdul SS. The Challenges, Problems and Strategies of Electronic Medical Record Implementation: a Case Study of an Eye Hospital from India. University of Tromso, Norway; 2008.
[7] Igira F, Titlestad O, Lungo J, Shaw V, Sheikh Y, Mahundi M, et al. Designing and Implementing Hospital Management Information Systems: Experiences from Zanzibar. Maputo, Mozambique: IIMC International Information Management Corporation; 2007.
[8] Heeks R. Information Systems and Developing Countries: Failure, Success, and Local Improvisations. The Information Society. 2002;2002(18):101–12.

Address for correspondence

Frank Verbeke, MD, PhD
Department of Biostatistics and Medical Informatics
Vrije Universiteit Brussel
Laarbeeklaan 103
B-1080 Brussels
Belgium
frank.verbeke@vub.ac.be

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-487

Implementing a National Scottish Digital Health & Wellbeing Service at Scale: A Qualitative Study of Stakeholders' Views

Ruth Agbakoba[a], Marilyn McGee-Lennon[b], Matt-Mouley Bouamrane[b], Nicholas Watson[a], Frances Mair[a]

[a] *Institute of Heath & Wellbeing, University of Glasgow, Scotland, UK*
[b] *School of Computing & Information Science, Strathclyde University, Scotland, UK*

Abstract

Digital technologies are being used as part of international efforts to revolutionize healthcare in order to meet increasing demands such as the rising burden of chronic disease and ageing populations. In Scotland there is a government push towards a national service (Living It Up) as a single point of reference where citizens can access information, products and services to support their health and wellbeing. The aim of the study is to examine implementation issues including the challenges or facilitators which can help to sustain this intervention. We gathered data in three ways: a) participant observation to gain an understanding of LiU (N=16); b) in-depth interviews (N=21) with stakeholders involved in the process; and c) analysis of documentary evidence about the progress of the implementation (N=45). Barriers included the need to "work at risk" due to delays in financing, inadequate infrastructure and skill-set deficiencies, whilst facilitators included trusted relationships, champions and a push towards normalisation. The findings suggest that a Scottish ehealth service is achievable but identifies key considerations for future large scale initiatives.

Keywords:
eHealth; Chronic Disease;Wellness Programs; Implementation

Introduction

Population ageing in the 21st century is a major issue with the World Health Organisation (WHO) forecasting that the number of people aged 60+ around the world is set to reach 2 billion by 2050 [1]. This represents the fastest growing age group anywhere in the world. While, this can be seen as a cause for celebration, ageing is changing the shape of society and therefore introduces enormous challenges, particularly in relation to the provision of health and social care services to a population with increasingly complex needs. A consequence of this demographic shift is the increased prevalence of non-communicable diseases (NCD) associated with ageing; also known as chronic diseases. Each year NCDs are the cause of 36 million deaths; in Scotland 60% of all deaths are attributable to a chronic condition and they account for 80% of all general practice consultations [2,3,4]. This has major implications for primary care provision within the United Kingdom (UK) as 90% of all patients' interaction within the National Health Service (NHS) 'starts and ends in primary care'[5]. Current models of care are unsustainable, costly and inadequate. There is a need for innovative approaches and solutions which can meet the demands of a fragmented system. The Scottish government aim to be at the forefront of innovation in becoming a "world class digital nation by 2020" with policy intending to help people to live longer and healthier lives at home or in a homely setting using digital technologies as an enabler [6]. The focus within primary care is on prevention, supported self-management and patient-centred holistic care. Healthcare is a lucrative and expanding market and the call to revolutionize it using digital technology has been seen as a key driver in creating innovation. However one of the most critical issues impeding previous efforts has been the gap between what we know can optimise health and wellbeing to what actually gets implemented in everyday practice. This has been referred to as a 'translational gap' where the normalisation of an intervention commonly fails [7]. The *purpose* of this study is to report on the *mid-point* views of stakeholders' on the factors which can promote or inhibit successful implementation of a large-scale digital health and wellbeing programme (Living It Up) across Scotland.

Materials and Methods

Participant Recruitment & Data Collection

In order to gain a wide range of perspectives and obtain a holistic picture of the implementation journey we contacted via email a purposive sample of stakeholders (N=16), representing local, national and international organisations. This sample spanned six sectors (industry, health and social care, housing, education, voluntary and statutory), all working together as a collaborative consortium within the Living It Up (LiU) project. Qualitative studies of stakeholders views are important to understand factors which affect implementations on the ground, for example a study on EHRs in Sweden identified discrepancies between the views of professionals and consumers which affected EHR utility and uptake of a national system designed to improve health provision [8].

Data Collection

Ethical approval for this study was obtained from the University of Glasgow, College of Medicine, Veterinary and Life Sciences Ethics Committee (2000130029). We collected and triangulated multiple sources of data through: a) prolonged participant observation; b) semi-structured interviews; and c) collection of a wide range of documentary evidence. Participant observation in this study involved two components: 1) observing monthly stakeholder meetings and 2) collecting data from quarterly meetings held between stakeholders (key informants) and researchers which served a primary purpose of capturing the changing face and shape of a digital health and wellbeing service which started as a 'blank

canvas'. These translated to a total of N=16 participant observation sessions over a period of 14 months (October 2012 to December 2013) and approximately 62 hours of interaction. Secondly, we conducted a total of N=21 semi–structured interviews (January to July 2014) which helped us identify and understand the barriers and facilitators to implementation. In order to capture both breadth and depth within the study we collected both a longitudinal and cross-sectional dataset. The longitudinal data included follow-up interviews conducted with N=6 participants (project managers) at baseline and mid-point (after 6 months). N=1 project manager resigned after 6 months and therefore we were not able to carry out their interview and this was not part of our final dataset. The cross-sectional dataset included mid-point interviews with N=10 participants that were service and technical partners involved in high-level strategic decision making. The interview questions were informed by Normalisation Process Theory (NPT) [9] and each interview lasted approximately 60 mins. All interviews were transcribed verbatim and coded in addition the research team undertook 'data coding clinics' where coding was discussed and agreed to ensure consistency and validity of the coding framework. Lastly, we collected N=45 multiple sources of documentary evidence such as Quarterly Progress Reports, Service Specifications, Alignment Interviews, Recruitment Reports and Evaluation Updates that were synthesised in order to capture and map out the implementation journey and to create a 'thick description' of the project over time. These methods maximized our ability to grasp the subjective behaviours of a multi-stakeholder environment.

Data Analysis

Each transcript was subject to theory-led qualitative analysis with reference being made to the Ritchie & Spencer (1994) thematic framework for data interpretation (familiarization; identifying a thematic framework; indexing; charting, mapping and interpretation)[10]. NPT was used as the theoretical framework underpinning our study as it is widely advocated and cited as a robust explanatory framework which captures the 'work' of implementing, embedding and integrating new technologies or services into routine practice [11]. The framework makes reference to four key domains namely: Coherence, Cognitive Participation, Collective Action and Reflexive Monitoring. The first domain looks at the '*sense-making'* work that people do individually or collectively in order to develop a shared understanding of a new intervention. The second domain reflects on the '*relational*' work that people do to encourage people to engage, buy-in and sustain a new intervention. The third domain simply refers to the '*operational work*' that people do and what needs to be done to ensure that the new intervention works in a real-life setting. The final domain looks at the '*appraisal*' work that people do to assess and understand the impact of a new intervention. We mapped our thematic findings to NPT in order to help us understand our data. The analytical process as a whole was facilitated using QRS Nvivo ® Version 10.0.

Results

The Makeup a Scottish Digital Health & Wellbeing Service

Living It Up (LiU) is part of a £37 million UK-wide project titled Delivering Assisted Living Lifestyles at Scale (dallas). The aim of dallas is to demonstrate how innovative technologies and services can be used for 'preventative care to promote independent living and improve peoples lifestyles' between June 2012 and May 2015 [12]. LiU is a digital platform (www.livingitup.org.uk) accessible via various modes of familiar technology, which aims to impact 55,000 people aged 50+ approximately 10% of the total Scottish population to improve their quality of life and independent living. The project targets 5 specific geographical locations namely; West Lothian, Moray, Highlands & Islands, Forth Valley and the Western Isles capturing a mix of urban, rural and remote rural areas. LiU has been hailed as a 'national ground breaking service' by government representatives. It aims to help the citizens of Scotland find local, trusted and personalised information on services which can support health and wellbeing [13]. The platform was developed using a co-design approach with intended users (members of the general public), creatives, technology designers and over 30 organisations (Figure 1). The LiU deployment is being led by the Scottish Centre for Telehealth and Telecare (SCTT) and NHS 24 which are government bodies' established with a purpose of facilitating the shift towards how health and social care services are provided, perceived and consumed.

Figure 1 –International Stakeholders for Living It Up

LiU is a platform that provides consumers with access to four key services: Connect, Discover, Flourish and Shine. The first being a service which supports *digital participation* among communities in providing a means for people to remain 'connected' with their friends and family as well as an opportunity to up-skill and learn how to go about using technology. This service enables users to remain 'connected' to their care-giver via Cisco Jabber Client video conferencing (VC) suite. The second service 'Discover' is based on *asset mapping* national and local information about organisations, services, activities and groups which consumers may find useful in meeting their needs. This service is powered by a national database called ALISS (A Local Information System for Scotland). This provides a personalised search and collaboration tool for users and enables organisations themselves to use it for sign-posting [13]. Users also have the ability to 'rate' services in an open format and share recommendations.

Flourish provides a suite of interactive tools to support people in *self-managing* their chronic condition. This includes approved health information and advice, text messaging alerts and remote-monitoring services to help support people with conditions such as Heart Failure (HF). The final service 'Shine' centres on community *capacity building*. It is advertised as the 'front door' to LiU. This service taps into the value of the contribution that citizens can make to society.

The service provides a 'profiling tool' which enables people to identify, nurture and refine their individual skills and experience in a way to 'give back' to their local community. This approach is being used to help contribute to improved wellbeing and stronger, more connected communities. Users of LiU can access the entire platform free of charge and there is an opportunity to become a member which will present them with a personalised dashboard. The final aspect of this platform is the 'Innovation Zone' which provides a space for enterprise where companies can advertise new solutions, apps or products which require testing or exposure ultimately fostering wealth creation [13].

Implementation Barriers

Working at Risk

This section provides an overview of the key mid-point themes that emerged as barriers during the national deployment of the LiU platform. A challenge amongst stakeholders in the beginning was to identify an agreed approach and direction. Due to the ambitious nature of the programme it was deemed important to engage with a wide range of different types of stakeholders. However, as this collaborative consortium combined a large number of organisations (local, national and international) from various backgrounds and with varying expertise, this introduced a degree of tension when voicing opinions, settling agreements and making decisions. This, coupled with the use of a co-design approach prolonged the design phases a great deal. For example, it was anticipated that a 'soft launch' of LiU would be live by March 2013 but in reality this occurred in November 2013 this attests to the scale of delay. Further to this, there were contractual difficulties which distorted the traditional tendering process and in order for stakeholders to proceed with their involvement in the project they usually had to 'work at risk' and obtain a letter of intent in relation to delayed financial payments. "*There was a huge bureaucratic delay in getting the thing set up and a contract out… The project was officially meant to start in June…you know, as a delivery partner, we did not get a contract until the following January. So, you know, we worked at risk. You know… to try and be a good partner but, you know, for a business that's not satisfactory and that means that you can't commit all the resources you would like to when you don't know if you're going to get a contract*". "*So, there's a slow start but what made it worse… was it took far too long to decide what LiU would be, you know…Our job was delivery of the requirement. Now, it took probably a year to decide on what the requirement was*"[LiU11]. "*LiU works in quarters. The financial approval process is so far off the pace of the work process that it's not only late it's almost at the end of the process…There's two things you can do as a supplier, one is you can say I'm not moving until I get approval…or else you can proceed at what's called…working at risk. I'd say it's uncommon*"[LiU20]

No Complete End-to-End Testing Environment

LiU aims to provide consumers with an integrated seamless journey of care. There was a consensus amongst stakeholders that delivering this vision was a complex process. Some partners worked in 'silos' within their organisations and concentrated on a particular piece of work which meant that difficulties were only recognised when that piece of work became integrated within the wider programme. This impacted the implementation process because there was no complete end to end testing environment across the entire programme. "*So basically what happens is a supplier, a technical partner is used to the principle of building something and then they put it into their test environment and test it. Right…now the problem is until you have interlinked them you don't know they are going to work. So supplier A could build product 1 and test it but when it goes live it might not work because of something that supplier B has running on their website…now it would cost a lot of money to build a complete end to end test environment"*[LiU20].

Inadequate Infrastructure – Challenging Boundaries

In some cases, the ability to deliver the innovation meant that transformational re-modelling of the current care model was required. This meant that LiU was being impeded in some aspects as current infrastructure was not suitable to adopt some elements of the platform."*I think we're certainly ahead of the game. Looking at international markets and speaking to our counterparts in the UK I think this is very much a pioneering project. ..We've actually moved to a kind of model that's maybe five years ahead of its time*". [LiU06]

Educating Stakeholders

In a multi-stakeholder environment, the need for all LiU consortium members to have a shared sense of understanding required a level of learning. It was clear that some stakeholders required more 'training' than others which slowed down the implementation process as well as the concept of innovation as a whole. "*There was a lack of understanding of digital technology and what it can do now. A lot of the people were not familiar with the use of digital technology and, you know, on the service side, the people that were designing the services did not themselves use this type of technology, so they were not pushing the boundaries*" [LiU11]

Designed for Local vs. International

This national platform aims to become Scotland's premier source for health and wellbeing; and stakeholders wanted to become a beacon for other countries but faced a challenge in identifying how to go about that. Several issues emerged in identifying the customer and market which led to concerns that the consortium might be taking too myopic a standpoint. "*The requirements were gathered from people in Scotland. Now, the market is not people in Scotland. The market is outside Scotland...for LiU to be commercialised…to become a product or service that people will buy…it needs to meet the needs of people outside Scotland. The current users are in Scotland but the future users are not in Scotland…""A big assumption was made that what suits Scottish people in the Scottish context will suit a world market and I think that's wrong". "There's a fundamental mismatch.* [LiU11].

Implementation Facilitators

Trusted Customer – Supplier Relationship

This section provides an overview of the key mid-point themes that emerged as positive enablers, or facilitators during the deployment of the LiU platform. The first being that over the course of the implementation, stakeholders developed a professional but friendly bond which changed the usual dynamics of the customer-supplier relationship. This introduced new ways of working in which representatives from sectors such as housing, healthcare and voluntary indicated that it helped drive the implementation forward. "*Normally that relationship is one of customer–supplier, and the public service has a very thorough obligation to treat all private organisations equally. You know, no favours, no special conditions and that's fine when you're trying to buy…you know…it's a plaster. It's just a question of who*

makes the cheapest plaster that passes the requirements". "But it's not a good way of handling things when you want to do innovation, because with innovation you need trust, you need a relationship, you need the ability to be able to say, in a trusted way from one side, this is what we want, and the other side says, well this is what, at the moment what I can deliver, but maybe I can move towards that over the next six months. And that's the only way that you can do joint innovation. And that's basically what Dallas has delivered…""We've moved from a customer–supplier relationship to a more of a partner relationship. And I think that's absolutely essential to solving some of these problems that we have in using technology to provide health care". "So I don't want to just be the telehealth guy. I'd hope we can be broader than that…""I'd hope we don't go back to customer- supplier". [LiU08]

Iterative User Feedback Shapes Development

Consumers have a crucial role as a stakeholder within the consortium as they are continuously consulted throughout the life cycle of the project in various ways both online and offline; and this provided an opportunity for grass-root level engagement and innovation to occur such as personalisation of health and wellbeing services. "*Sometimes they'll be giving the feedback on a one-to-one basis at workshop events and that goes through…we consolidate that…to shape the development of LiU. If they go through the digital portal then…that goes directly to support office who then again push that out …and see if things can be improved* "[LiU05]

Local Champions Driving Implementation

A key facilitator has been the establishment of local champions who are people that live in the target communities that either a) have a vested interest in co-designing LiU or b) have identified the value of LiU as part of their daily lives. They have been identified as a key driving force in creating awareness and encouraging regular people to buy-in to LiU. *What I've done is I've been very lucky and I've got a great group, a core group of local community champions, who are basically... in a way, I'm leaving it to them, because I think it sells it better if it's coming from actual users. So we've got one guy who's writing a regular blog about living with long-term conditions, you know, he gives practical advice based on his own experiences, and that's been very popular with people, given it a human edge, if you like*. [LiU07]

Product Ownership & Business Opportunities

It was largely agreed by small and medium-sized businesses (SMBs) that working on a large national collaborative project such as LiU created new business opportunities and ventures. This helped to provide a platform in which the vision of wealth creation and innovation could be achieved. More importantly stakeholders were identified as product owners in different elements of LiU and therefore this enabled them to showcase this work as well as a 'collection' of individual works within their respective organisations; an opportunity that they may not have had without being part of LiU. "*As a company we've had great benefit from being a part of this project because it has allowed us to establish a position in the individual health market and you know, we're working for Living It Up, we're working for all the Dallas projects. So, we're not restricted to Living It Up, although that's given us opportunity*" [LiU11]

Push Towards Sustainability and Normalisation

The push towards scaling LiU and making it sustainable far beyond the official end date of 2015 has positively influenced the implementation process. Stakeholders themselves are thinking long term but more importantly in 'real-terms' as to how LiU can be integrated into daily practices. This is a key overall positive factor in ensuring that the project as a whole is a success. "*We're going to run some workshops, actually, just to see how health and care professionals can implement if there's some of the tools. I mean, not all of them, but we're going to have just some of the tools that are relevant to them and their clients or their service users. So we will be kind of running workshops in all of the areas just on how we can do that, and actually just get them to implement it in as part of their daily working*" [LiU_15]

Discussion & Conclusion

This study explored the views, knowledge and understanding of stakeholder personnel and organisations involved in the deployment of an on-going national digital health and wellbeing project at scale in Scotland. The results of the study show that obtaining stakeholders views on factors affecting the implementation process provides valuable insights which can help to inform its future development in becoming a sustainable service for Scottish citizens. A limitation of our work is the lack of data from end-users of the LiU services and the fact that we are describing a deployment still "in process", however a strength of this study has been the use of the NPT framework in capturing the 'work of implementation' as well as providing a basis for learning and critical reflection in understanding the valuable lessons that have been learned throughout this journey of implementation. The use of the NPT framework has helped us to highlight barriers and facilitators and we apply it here in order to interpret and synthesise our findings.

Coherence

This domain refers to the '*sense-making'* work that people do individually or collectively in order to develop a shared understanding of a new intervention. It is clear that there was some difficulty experienced by stakeholders in developing a shared direction of travel due to several factors such as the number of stakeholders involved in the process, identifying requirements to match future need and having to 'work at risk'. Although facilitators such as the creation of 'trusted' relationships and the move towards embedding this intervention into everyday practice has helped to overcome this barrier. It seems that these risks were necessary when implementing a project which is at the forefront of innovation. Recent research has confirmed that having good existing relationships or links between senior management or strategic level players helps to improve communication among implementers as well as securing long lasting change [14]. Our findings within this domain clearly demonstrate the need to understand organisational cultures as a key ingredient and basis for any innovative digital health and wellbeing project.

Cognitive Participation

This domain refers to the '*relational*' work that people do to encourage people to engage, buy-in and sustain a new intervention. A key barrier that needed to be addressed was lack of knowledge/skill-set deficiencies and the need to educate, upskill and train stakeholders in digital technologies as it was clear that not all stakeholders had the same level of understanding. This finding unearthed a link between the 'Collective Action' domain and the 'Cognitive Participation' domain due to the fact that this process was required to take place before stakeholders actually engaged with LiU in order for them to go on to endorse or promote it themselves. Local champions however helped to overcome the barriers that

stakeholders faced by not concentrating on the technology but on personalising the benefits to demonstrate to potential users how this product can help them in their daily lives. Previous research in Australia has shown that use of clinical champions can play a critical role in helping to promote uptake and sustainability of telehealth; with the authors pointing out that it is more important to get the service model right rather than the technology itself [15]. This is key as there is a lack of evidence in relation to participation and engagement within the field of ehealth and wellness [16].

Collective Action

The third domain refers to the '*operational work*' that people do and what specifically needs to be done to ensure that a new intervention works in a real-life setting. Barriers which affected the practical application of LiU included inadequate infrastructure, constraints on resources (including finances) and limited testing environments which are key findings that align with available evidence from the United States of America [17]. In Scotland there remains a challenge in delivering services to people living in remote locations which compounds the existing burden on the system. The need for adequate infrastructure and resources to support digitally enabled self-care has been recognised and the Scottish Government have recently launched a national programme to enhance the current broadband and fibre optic capabilities.

Reflexive Monitoring

The final domain looks at the '*appraisal*' work that people do to assess and understand the impact of a new intervention. A challenge was designing to meet all needs but positive themes such as the creation of business opportunities and iterative user feedback emerged as key facilitators in assessing the impact of LiU. Particular focus on the latter finding is significant as it illustrates the value of input from Scottish citizens in dictating their own care and becoming 'active' recipients with increased choice about how and where they receive services as opposed to the traditional passive role that is played. There is a considerable amount of value from capturing the process and journey of implementation at scale. Lessons that have emerged as key learning points include the need for flexible and trusted working environments to support multi-sector working partnerships and the need for policy to support innovative business models. This report highlights difficulties faced in delivering new digital health and wellbeing services at scale and the need for further research to help understand implementation issues in order to a) bridge the 'translational gap' and b) inform future ehealth policy and practice.

Acknowledgments

We would like to thank the Scottish Government and Innovate UK (Technology Strategy Board) who are sponsoring the Living It Up project and the Medical Research Council UK who are funding our research.

References

[1] World Health Organisation. WHO | 10 facts on ageing and the life course [Internet]. 2014 [cited 08 November 2014]. Available from: http://www.who.int/features/factfiles/ageing/en/

[2] World Health Organisation. WHO | Noncommunicable diseases [Internet]. 2014 [cited 11November 2014]. Available from: http://www.who.int/mediacentre/factsheets/fs355/en/

[3] Scotland.gov.uk. Long Term Conditions [Internet]. 2014 [cited 11 November 2014]. Available from: http://www.scotland.gov.uk/Topics/Health/Services/Long-Term-Conditions

[4] Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012 Jul 7;380(9836):37-43

[5] Transforming primary care in London. NHS England (London Region)/Primary Care Transformation Programme. 2013.

[6] Digital Participation A National Framework for Local Action [Internet]. Scottish Government; 2014 [cited 15 November 2014]. Available from: http://www.scotland.gov.uk/Resource/0044/00448804.pdf

[7] Morrison D, Mair F. Telehealth in practice: using Normalisation Process Theory to bridge the translational gap. Primary Care Respiratory Journal. 2011;20(4):351.

[8] Lehnbom EC, McLachlan AJ, Brien JA. A qualitative study of Swedes' opinions about shared electronic health records. Stud Health Technol Inform. 2013;192:3-7

[9] May C, Finch T. Implementation, embedding, and integration: an outline of normalization process theory. Sociology 2009;43:535-54

[10] Ritchie J, Lewis J. Qualitative Research Practice: A Guide for Social Science Students and Researchers. London, england: Sage Publications; 2014.

[11] McEvoy, Ballini L, Maltoni S, O'Donnell CA, Mair FS, MacFarlane A. A qualitative systematic review of studies using the Normalisation Process Theory to research implementation processes. Implementation Science 2013, 9 (2).

[12]McGee-Lennon, M.; Bouamrane, M-M.; Barry, S., Grieve, E.; Latina, D.;Watson, N.;O'Donnell, K.; Wyke, S.; Brewster, S.; Briggs, A.; Finch, T.; Mair, F. Evaluating the Delivery of Assisted Living Lifestyles at Scale (dallas). Proceedings of HCI 2012 - People & Computers XXVI, The 26th BCS Conference on Human Computer Interaction, Birmingham, UK, Sept. 2012

[13] Living It Up Progress Report [Internet]. NHS 24; 2014 [cited 5 December 2014]. Available from: http://www.nhs24.com/aboutus/nhs24board/agendasandpapers/2014/april/

[14] Ling T, Brereton L, Conklin A, Newbould J, Roland M. Barriers and facilitators to integrating care: experiences from the English Integrated Care Pilots. International Journal of Integrated Care 2012;12:e129.

[15] Wade VA, Eliott, JA and Hiller JE. Clinician acceptance is the key factor for sustainable telehealth services. Qualitative Health Research 2014; 5: 682-694

[16] Mair, F., May, C., O'Donnell, C., Finch, T., Sullivan, F., Murray, E. (2012). Factors that promote or inhibit the implementation of e-health systems: An explanatory systematic review. Bulletin World Health Organisation,90(5), 357-364.

[17] Goroll AH, Simon SR, Tripathi M, et al. Community-wide implementation of health information technology: the Massachusetts eHealth Collaborative experience. J Am Med Inform Assoc 2009 Jan-Feb;16(1):132-9.

Address for correspondence

Ruth Agbakoba, 124 Observatory Road, G12 9LX, Scotland, U.K.
Email: r.agbakoba.1@research.gla.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-492

Evaluating a Proof-of-Concept Approach of the German Health Telematics Infrastructure in the Context of Discharge Management

Ursula Hübner[a], Georg Schulte[ab], Björn Sellemann[c], Matthias Quade[c], Thorsten Rottmann[c], Matthias Fenske[d], Nicole Egbert[a], Raik Kuhlisch[e], Otto Rienhoff[c]

[a] *Health Informatics Research Group, Hochschule Osnabrück, Osnabrück Germany*
[b] *Klinikum Osnabrück, Osnabrück Germany*
[c] *Department of Medical Informatics, University Medical Centre, Göttingen Germany*
[d] *Diakoniewerk Osnabrück, Osnabrück Germany*
[e] *Fraunhofer FOKUS, Berlin Germany*

Abstract

Although national eHealth strategies have existed now for more than a decade in many countries, they have been implemented with varying success. In Germany, the eHealth strategy so far has resulted in a roll out of electronic health cards for all citizens in the statutory health insurance, but in no clinically meaningful IT-applications. The aim of this study was to test the technical and organisation feasibility, usability, and utility of an eDischarge application embedded into a laboratory Health Telematics Infrastructure (TI). The tests embraced the exchange of eDischarge summaries based on the multiprofessional HL7 eNursing Summary standard between a municipal hospital and a nursing home. All in all, 36 transmissions of electronic discharge documents took place. They demonstrated the technical-organisation feasibility and resulted in moderate usability ratings. A comparison between eDischarge and paper-based summaries hinted at higher ratings of utility and information completeness for eDischarges. Despite problems with handling the electronic health card, the proof-of-concept for the first clinically meaningful IT-application in the German Health TI could be regarded as successful.

Keywords:

eHealth Strategy; Health Telematics; Electronic Health Card; Nursing; Patient Discharge.

Introduction

National eHealth and Health Telematics strategies came into existence more than a decade ago [1] and have been implemented with varying success [2, 3]. In Germany, the Health Telematics Infrastructure (TI) based on an electronic health card for the patient and an electronic health professional card, was announced in 2004 but was delayed several times due to politically motivated criticism of the physician associations and other stakeholders [4]. Consequently, systems based on other technologies, and pursuing other approaches, were developed [5, 6]. In the meantime, the development of the Health Telematics Infrastructure (TI) had been carried on, the use of the electronic health card all over Germany is mandatory from January 2015 on [7], and basic IT-applications such as the online management of patient demographics are ready for piloting in several federal states. So far the electronic health card contains administrative data only. The clinical eHealth-applications that had a true chance to run on the Health TI were rather unclear at the onset of the project, so was a realistic timetable of the availability of the TI.

Against this background, it seemed desirable to develop and test applications with a high potential impact on patient care that make use of TI elements such as the electronic health card as a gatekeeper to the patient data. Among these applications, the electronic health record possesses paramount importance for supporting all health care processes as well as for supporting the discharge process. Discharge management is particularly crucial in patients with chronic conditions and is reported to be problematic in many countries [8]. However, when organised properly, discharge management can help significantly. This, for example, could be demonstrated in geriatric patients [9], just to mention one group of patients that benefit from these measures. Elderly patients often require medical and nursing information to be transmitted at discharge to the institution providing follow-up care in order to maintain the continuity of care [10].

We, therefore, aimed at developing a prototype for an eDischarge application based on an EHR that was embedded into a laboratory Health TI and served as a proof-of-concept for the TI. The eDischarge application should transport administrative, nursing, social and medical data between health care organisations. In this study, we particularly wanted to evaluate its technical-organisational feasibility, usability, information completeness and utility.

Materials and Methods

Discharge application embedded into Health TI

The eDischarge application was built on the ground of the German national HL7 CDA based standard of the eNursing Summary [11]. It included data about the patient and the health care professionals involved in the structured CDA header, and nursing, social and medical data in the body. In detail, the body contained the nursing process, social information, references to legal documents, home care status and medical information, mainly medical diagnoses and medication (extracted with authorisation from the medical summary). The eDischarge solution was a Web application that was connected to a central electronic health record (EHR), in which the eNursing Summary documents were stored. The eDischarge solution also included a certified card reader for the electronic health card (eHC). Patient's demographics were

read via the eHC and nursing, social and medical information of the patient were entered manually via the eDischarge client. Authentication via the health professional card was simulated by a software-key. Patient data could only be accessed when this key was invoked and the eHC was inserted (authorisation and authentication principle of the Health TI). As the current version of the eHC card that had been rolled out to the citizens did not have a pin-code on it, a second card carrying this information was used in the tests.

Certified card readers for the eHCs, software-key, electronic health record kernel and server and secure client-server communication constituted the laboratory Health Telematics Infrastructure, which had been developed by Fraunhofer Fokus in a previous project and was extended in this project to demonstrate the feasibility of implementing the Health TI specifications. The eDischarge application and the laboratory Health TI formed the proof-of-concept to be evaluated. The EHR server was located at the University Hospital Göttingen to comply with the privacy and data security regulations.

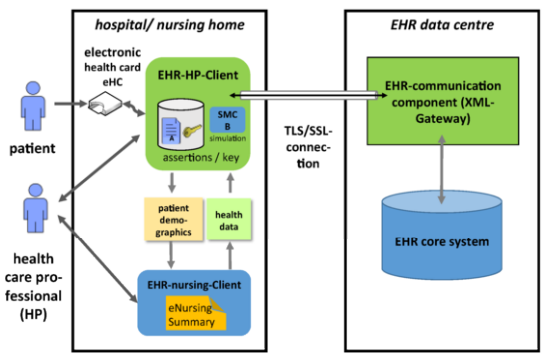


Figure 1 – Laboratory Health Telematics Infrastructure (TI)

Evaluation phases

The eDischarge application and its communication with the EHR server were evaluated during four phases, in which we aimed at measuring different targets (Table 1). The applications tested comprehended the creation and update of the eNursing summary and its storage into and retrieval from the central EHR server. Thus "sending" and "receiving" a discharge document were technically identical to storing and retrieving it centrally.

Table 1 – Evaluation Phases and their duration

No.	Primary aim	Duration
Ia	technical feasibility	(14/01 – 03/08/2014)
Ib	technical-organisational feasibility	(04/08 – 31/08/2014)
II	usability	(01/09 – 28/09/2014)
III	usability and utility	(29/09 – 31/03/2015)

In phase Ia, we wanted to know what types of problems occurred once the application was transferred from the technical laboratory of the developers to the *field laboratory* of the evaluators. The system was operated by two nurses of the evaluation team. In phase Ib, the evaluation took place at two clinical sites with two chief users (nurses) testing the system in a *realistic environment (office at clinical site)* with 20 realistic but fictitious discharge documents. Phases Ia and Ib covered the issues of a formative evaluation where the system still was technically adapted in accordance with the evaluation results. During all the phases, the users recorded the type, location and severity of the problems in standardised and structured log books. In addition, usability was assessed in phases II and III by the IsoMetrics questionnaire [12]. During these two phases, nurses, who were on duty, operated the system in the *same realistic environment* as in phase Ib. In phase II, data of six past discharges (real documents with modified identification data) were sent. Finally in phase III, 13 ongoing discharges were included in the study. Phase III was particularly designed to measure information completeness and utility of the eDischarge application in comparison to a paper-based discharge. To this end, discharge patients were randomly allocated to the eDischarge group or the paper discharge group. In phase III, we were also interested whether there were specific problems using the electronic health card and obtaining the declaration of consent.

Evaluation setting

We chose two evaluation sites in Osnabrück that were used to cooperating and communicating with each other; a municipal hospital with 717 beds (Klinikum Osnabrück) and a health system with nine nursing homes and three ambulant nursing services (Diakoniewerk Osnabrück). The testing was based on bi-directional communication, thus we looked at the discharge from the hospital to the nursing home and also from the nursing home to the hospital. To this end, we involved the Department of Neurology at Klinikum Osnabrück and Küpper-Menke-Stift (137 residents) as one of the nursing homes within the Diakonie. Two chief users from Klinikum Osnabrück and Diakonie accompanied the study in phases Ib to III.

During two information meetings, representatives from the two evaluation sites were informed about the study, its aims, the procedure and the technology involved. There were four user trainings that took place on the 21st and 25th of July 2014 in Küpper-Menke-Stift and the 22nd and 24th of July 2014 in Klinikum Osnabrück. In case of problems, the two chief users supported the users when operating the system. All users belonged to the group of nurses. The patients involved in phase III testing had to give their informed consent to take part in the study and also to allow the organisations to transfer their discharge data to the other organisation. The study design obtained approval from the ethics committee of Klinikum Osnabrück.

Pretest eNursing Summary as the eDischarge data set

In order to make sure that the HL7 eNursing Summary standard was suitable for the eDischarge application, we tested it against commonly used discharge forms. We were particularly interested in whether there were any data in the forms that were not covered by the HL7 standard and vice versa. We drew a regionally clustered random sample of 375 institutions throughout Germany (hospitals, ambulatory nursing services and nursing homes), which covered all three types equally. They were asked to send us their discharge forms. We received a set of 69 forms, which was enriched by forms retrieved from the Internet. Finally, a set of 114 different forms was obtained and independently analysed by two nurses regarding the match of potential data entries.

Results

Pretest results

The analysis of the 114 forms showed that all nursing, social and medical information found in the forms could be matched with fields in the HL7 standard. Only very specific information that was found in individual forms that concerned organisational issues and specific information on hygiene and physiotherapy could not be represented in the standard. In contrast, none of the forms contained all information items of the HL7 standard. Most of the forms were problem-oriented (77%). However, even if the problem could be entered, only in 23% the reason could be inserted. Only in 45% the interventions, and in 44% the means that were associated with a problem could be given. We concluded that the HL7 eNursing Summary was a valid and suitable standard for the eDischarge application, which contained more information to be carried than the paper forms.

Technical-organisational feasibility

The evaluation results are presented in the following in accordance with the respective research questions and topics.

In phase Ia, a series of many problems occurred. Examples are: installation of the clients in the hospital and the nursing home, the compatibility of the card reader, moving the server to the University Medical Centre Göttingen, terminology used in the application, workflow and pdf presentation. These problems could be solved during a continuous formative evaluation with several feedback loops that took eight months all in all (Table 1). The actual evaluation started in Phase Ib. In the following the results, which were obtained from analysing the logbooks in terms of the number and severity of problems, are presented. The logbooks allowed entries describing the following steps of operating the application: starting the program, reading patient demographics, creating a new summary (inserting, buffering, resuming, changing and deleting), transforming into pdf, printing, storing in EHR (sending), opening/browsing EHR and retrieving/viewing eDischarge summary (receiving). Problems and their severity could be documented for each of these steps.

The percentage of problems per discharge session (Figure 2) decreased from phases Ib to II and rose again when operating the system during real discharges (phase III).

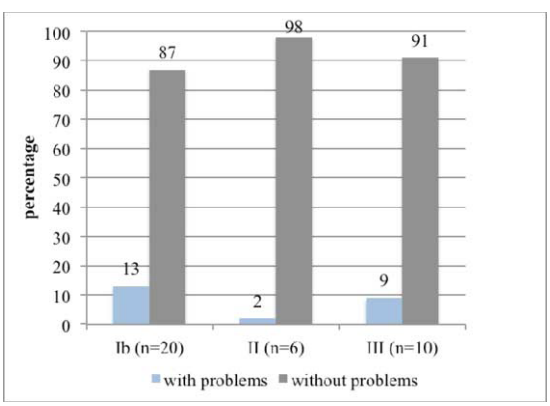


Figure 2 – Mean percentage of problems in phases Ib – III

Out of the 26 individual problems in phase Ib, 46% were rated as "1" and 46% as "3". On a severity scale from 1-3, "1" was denoted as "a problem that could be solved during the session" and "3" as "several problems that could not be solved during the session". Four percent of the problems in Ib were rated as "2" and 4% as "4". In phase II there was only one problem that was judged "1" and in phase III 25% of the 12 problems received the lowest severity level "1" and 75% level "3".

Usability

Usability data from phases II and III were pooled due to the small number of different nurses who actually used the system. Selected results for suitability of the task (median 4), error tolerance (median 3.25) and suitability for learning (median 3) of the IsoMetrics questionnaire [12] are shown in Figure 3. The scale ranged from 1 (strongly disagree) to 5 (strongly agree).

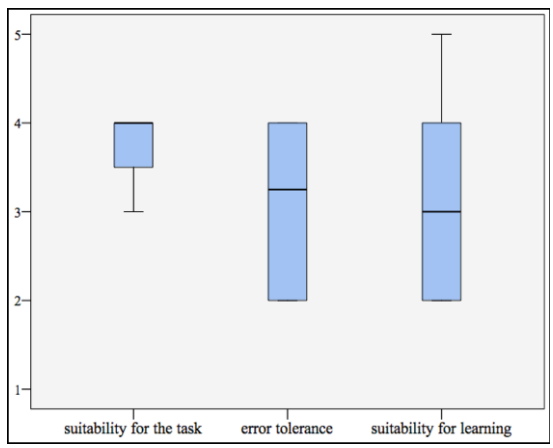


Figure 3 – Usability profile in phases II and III (n = 6 nurses)

eDischarge vs. paper discharge: utility and completeness

Fourteen patients met the criteria for inclusion in phase III. In one case, no consent for the participation could be obtained. In the other case, the electronic discharge summary could not be opened by the recipient due to a severe error (type 3). Thus, nine electronic and three paper discharges were rated with on the utility and completeness by the recipient on a six-point Likert scale (1 being very good and 6 very poor) in phase III.

Electronic discharges were rated better than paper discharges both with regard to utility and completeness (Figure 4). This judgement corresponded with the sum of entries of 18.4 (±2.5) for eDischarge summaries and 10.7 (±6.7) for paper summaries. Table 2 shows the average number of entries per category.

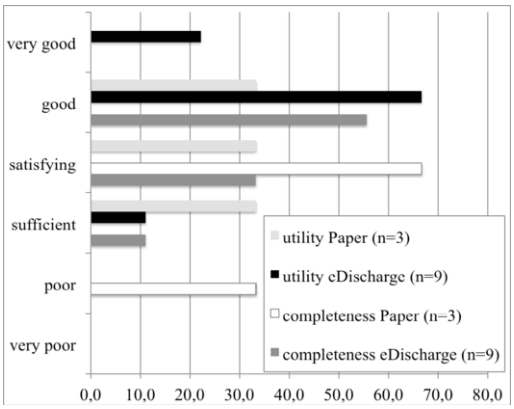


Figure 4 – Ratings of utility and completeness by recipients

Table 2 – Average number of entries

Category	eDischarge n=9	Paper discharge n=3
sender	1	0.3
recipient	1	1
patient demographics	4.4	2.7
nursing problem/intervention	4.3	2.7
assessments	1.0	0
social/administrative information	4.3	2.0
living environment	1.6	0.7
comments	0.8	1.3

The number of entries in eDischarge summaries exceeded the number in paper discharges in all categories with the only exception of "comments" and information about the "recipient". There were no assessments mentioned in the paper-based discharge form because the nurses decided not to include it in the summary.

There were a couple of problems associated with the handling of the electronic health card; in one case, the card was not available, and in the other case, the patient belonged to the group of privately insured persons who are not obliged by law to obtain an electronic health card. There was a general problem of getting the informed consent for sending the data electronically to the other institution. All the patients involved could not decide for themselves and thus the consent from a relative or the guardian had to be given. This procedure turned out to be complicated.

Discussion

This study is the first to investigate the feasibility and usefulness of the German Health Telematics Infrastructure in a clinically meaningful scenario. In addition, it is the first to evaluate the TI using a German national HL7 based standard, the eNursing Summary CDA. It is noteworthy, that the scenario tested mainly embraced nursing discharge management, which is a scenario that has been severely neglected by German TI executives before [13] and that still is not covered properly by the law [14].

There was a rather long period that was dedicated to formative evaluation in phase Ia with many feedback loops and meetings between the developers and the evaluation team. This long time proved to be absolutely necessary for having a system that worked in an acceptably stable manner and was adjusted to the workflows both at the hospital and the nursing home. The evaluation would not have been possible if the system had been taken out of the technical laboratory into the field without modifications. This particularly held true because the application still was in a proof-of-concept state. A dedicated phase of intertwined less formal evaluation and technical adaptations therefore seems advisable for applications in early stages to be evaluated successfully later.

The results of the formal evaluations in the following phases show that the software application stabilised with regard to the number and the severity of problems from phase Ib to II reaching almost no problems at all. However, the number and severity rose again in phase III, which was probably caused by the new users operating the system. They had received a training session but still may have felt unsure about how to use the system correctly.

The most important finding with regard to the usability was the good and quite unanimous rating for the suitability for the task. Analysing the usability also showed a median of 3.25 for error tolerance, which is slightly above the value in the middle of the codomain and corresponds with the number of errors particularly in phase III. Also, the suitability for learning was rated only moderately with a high variability showing that there is still enough room for improvements.

What counts at the bottom line is what the receivers say. Their judgement about the utility and the completeness gives an account of what the system can achieve. In both cases, the eDischarge approach yielded better ratings than the paper discharge. This was not surprising. It was very likely caused by the different types of forms used for eDischarge and for discharge as usual, i.e., a two-page paper form designed by Klinikum Osnabrück and a one-page summary used by Küpper-Menke-Stift. Whereas the eNursing Summary allowed the user to insert many different types of information and as often as necessary, the paper form was restricted to the maximum of two pages and so was the amount of information.

All of these results have to be treated with caution due to the small number of discharges in phase III. More data will have to be collected.

There are several limitations due to the laboratory Health TI into which the eDischarge application was embedded. They include first and foremost the complicated workaround because of the missing pin codes on the current version of the electronic health card and the non-usage of health professional cards for nurses. Their implementation is planned but has not yet materialised.

Adoption of IT standards such as the HL7 CDA based eNursing Summary and the eMedical Summary are rather slow in Germany due to unclear regulations. This was the reason why we could not test the standard in a fully fledged manner, i.e., connecting systems from different vendors.

Problems with handling the eHC that had been anticipated generally in the context of the Health TI actually occurred in

our small sample and therefore seem to be quite likely. Card based access control is a safe and elegant way but can cause a considerable delay in the process once they are not available. In this test environment, we could overcome the missing eHC via the manual input of patient data and via the second card with the pin code that was produced for each patient in this test. In the real world scenario, a missing card would have prevented the health care professional from reading the data not only on the card but also in the EHR. Thus, patients controlling their own health data is a highly desirable goal but it can turn out to be a big barrier when the card and the patient have to be at the same place at the same time. It is therefore advisable to rethink the use of eHCs in a discharge scenario.

Obtaining informed consent for elderly people who need assistance due to physical or mental disorders was not easy, but is feasible. Regulations for allowing health data to be transmitted electronically between health organisations in Germany, therefore, must be changed. A more generic solution seems appropriate, i.e., a citizen giving an informed consent that applies to many cases of when the data need to be transferred.

Meanwhile, the German government announced an eHealth act to give a fresh impetus to the Health TI and its applications [15]. The act addresses various clinically meaningful scenarios, in particular the electronic discharge summary, the medication plan and the emergency data set. This underpins the importance of this study, which provides useful insight into the mechanism of one of the high priority scenarios. However, it needs to be said that the eHealth act in its current version refers to nurses only as potential future users, despite its focus on the demographic change.

In summary, this study could demonstrate the technical-organisational feasibility of Health TI supported discharges. The usability of the web-based eDischarge application definitely needs to be improved for daily routine usage and to be integrated into the real Health TI. The study also highlights the importance of electronic data and electronic data transmission in terms of completeness of the information and demonstrates the utility of eDischarges. It thus makes the case for eDischarges in general and in nursing particularly. eNursing summaries and their standardisation and implementation must become a high priority goal in eHealth, and this not only in Germany [16].

Acknowledgements

This study has been funded by a grant of the German Federal Ministry of Health. We would like to warmly thank all participants (staff and patients) from the two organisations for their consent, time and effort.

References

[1] Hübner U. European Health Telematics. In: Ball MJ, Douglas JV, Hinton Walker P, DuLong D, Gugerty B, Hannah KJ, Kiel J, Newbold SK, Sensmeier J, Skiba DJ, Troseth M (eds.) Nursing Informatics: Where Caring and Technology Meet. 4th ed. London: Springer, 2011; pp. 375-400.

[2] Greenhalgh T, Morris L, Wyatt JC, Thomas G, Gunning K. Introducing a nationally shared electronic patient record: case study comparison of Scotland, England, Wales and Northern Ireland. Int J Med Inform. 2013;82(5):e125-38.

[3] Winblad I, Hämäläinen P, Reponen J. What is found positive in healthcare information and communication technology implementation?-the results of a nationwide survey in Finland. Telemed J E Health. 2011 Mar;17(2):118-23.

[4] Lang A, Mertes A. Introduction of the electronic health card in Germany: influence of interest positions and sector membership on the establishment of an implementation network. Gesundheitswesen. 2011 Jan;73(1):e12-20.

[5] Elektronische Fallakte. http://www.fallakte.de/

[6] N. Gusew, T. Bartkiewicz, W. Bautsch, A. Gerlach, M. Goldapp, R. Haux, U. Heller, H. P. Kierdorf, T. Kleinschmidt, W. Ludwig, U. Markurth, S. Pfingsten-Würzburg, M. Plischke, H. Reilmann, R. Schubert, C. Seidel, R. Warnke. A Regional Health Care Network: eHealth.Braunschweig. Domain Fields and Architectural Challenges. Methods Inf Med 2012 51 3: 199-209.

[7] Krüger-Brand H. Gesundheitskarte: Pflicht ab 1. Januar 2015. Dtsch Arztebl 2014; 111(35-36): A-1444 / B-1249 / C-1189.

[8] Schoen C, Osborn R, How SK, Doty MM, Peugh J. n chronic condition: experiences of patients with complex health care needs, in eight countries, 2008.Health Aff (Millwood). 2009;28(1):w1-16.

[9] Steeman E, Moons P, Milisen K, De Bal N, De Geest S, De Froidmont C, Tellier V, Gosset C, Abraham I. Implementation of discharge management for geriatric patients at risk of readmission or institutionalization. Int J Qual Health Care. 2006;18(5):352-8.

[10]Hübner U, Flemming D, Heitmann KU, Oemig F, Thun S, Dickerson A, Veenstra M. The Need for Standardised Documents in Continuity of Care: Results of Standardising the eNursing Summary. Stud Health Technol Inform. 2010;160: 1169-73.

[11] HL7 Germany. ePflegebericht. http://wiki.hl7.de.

[12]Gediga G, Hamborg KC, Düntsch I. The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. Behaviour and Information Technology, 1999;18(3):151 - 164.

[13]Hübner U. Telematik und Pflege: gewährleistet die elektronische Gesundheitskarte (eGK) eine verbesserte Versorgung für pflegebedürftige Bürgerinnen und Bürger? GMS Med Inform Biom Epidemiol. 2006;2(1):Doc01.

[14]Sozialgesetzbuch V (SGB V) §291a

[15]Federal Ministry of Health. Referentenentwurf eines Gesetzes für sichere digitale Kommunikation und Anwendungen im Gesundheitswesen (eHealth act) 13th Jan. 2015.

[16]Hübner U, Saranto K, Coenen A, Sensmeier J. eNursing Summary – Where Global Standardisation and Regional Practice Meet (Panel). In: Lehmann CU, Ammenwerth E, Nohr C (eds.) MEDINFO 2013 Proceedings. 2013; p. 1236.

Address for correspondence

Prof. Dr. Ursula Hübner Hochschule Osnabrück Health Informatics Research Group, D-49009 Osnabrück, Germany u.huebner@hs-osnabrueck.de, www.wiso.hs-osnabrueck.de/igw_team.html

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-497

Monitoring Telemedicine Implementation in Denmark

Christian Nøhr[a], Sidsel Villumsen[a], Stephanie Bernth Ahrenkiel[b], Lars Hulbæk[b]

[a] *Department of Development and Planning, Aalborg University, Denmark*
[b] *MedCom, Denmark*

Abstract

According to the literature, Denmark has a leading position in the dissemination and use of health informatics. However, there is a lack of systematically collected and documented experience of telemedicine solutions in Denmark. This evidence is being established with a new project, which maps out all telemedicine initiatives in Denmark. Data on all the initiatives is collected in a single database and some of the data is analyzed in this paper. It is shown that there are a very large number of telemedicine initiatives in Denmark and that the elements from the national strategy for telemedicine are clearly visible in the telemedicine map. The very large number of projects could, however, also indicate a lack of national coordination of initiatives and a need for evaluation activities to systematically collect and communicate the learning outcomes from all the new projects.

Keywords:

Telemedicine; dissemination; monitoring; evaluation.

Introduction

Denmark is among the world leaders in digitization of the health care sector [1, 2]. The small size of the country (5.5 mill. inhabitants), the relatively well-organized health sector, which is almost entirely publicly financed by tax money and the fact that every citizen has had a unique personal identifier since 1968 has enabled an efficient and penetrating implementation of computer systems. The systems were in the beginning mainly used for administrative purposes, but during the last decades also for clinical applications. All hospitals have EHR systems, including clinical notes, medication, all kinds of laboratory data, and imaging RIS/PACS. All general practice physicians are fully digitized, and citizens can all view their medical record online including lab tests and medication, renew prescriptions and in many cases book consultation with their GP. Because of the unique personal identifier and a pioneering work with the development of a national health data network, communication of health data has also been possible for a long time, which means that clinicians use it routinely. Of course the development of the many different health information systems and their integration has not been a walk in the park, many problems have had to be overcome and many conflicts have been resolved – it rather means that there exists a huge amount of experience.

Due to the pioneering efforts in health informatics, huge efforts have also been exercised in telemedicine, where Denmark also has been mentioned as the world leader by several sources [3, 4]. The Danish Government, Local Government Denmark, and the Danish Regions launched a national action plan for dissemination of telemedicine in August 2012 [5]. The plan outlines a number of new possibilities with telemedicine:

- Improved and more coherent patient care
- More individually planned treatment and self-reliant patients
- Increased professional competences across sectors
- Financial gains in municipalities and regions

To realize these potentials, a stronger framework must be established which includes reference architectures and standards, common assessment models, and overview of telemedicine solutions.

The national action plan for telemedicine notes that there is a lack of systematically collected and documented experience of telemedicine large-scale solutions that have been in operation over a longer period of time. To gain such experience, five specific telemedicine initiatives have been launched: clinically integrated home monitoring; home monitoring for COPD patients; tele-psychiatry; internet psychiatry; and national telemedical assessment of ulcers.

To produce an overview of telemedicine solutions and as a first step to systematize the wealth of telemedicine experience, all existing telemedicine projects in Denmark have been mapped in a single database [6]. The objective of the database is to annually collect and publish an overview of the diffusion of telemedicine in health care. The mapping is updated continuously through a proactive outreach work from MedCom. For each telemedicine project, the database contains data in the following categories:

- Master data (title, aim, etc.)
- Involved actors (managers, users, etc)
- Disease area
- Activities (Consultation, diagnostics, screening, monitoring, shared care, etc.)
- Relation to specific trajectory programs
- Applied technologies (hardware, software, specific integration to other systems)

The inclusion criteria of the database has been the WHO definition of telemedicine: "*The delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities*" [7]. The database contains data of 372 projects, which have been collected among national initiatives, initiatives in the five regions and the 99 municipalities in Denmark. As the private healthcare sector in Denmark with regards to health informatics and telemedicine is negligible, the coverage of the existing projects is regarded complete.

The database provides an opportunity to monitor the actual state of telemedicine in Denmark. The aim of this study is to

explore the numerous telemedicine projects in operation in Denmark with a particular focus on the aim of the telemedicine projects, the actors involved, the specific activities and the applied technology.

Methods

The study is designed as a cross section study. The data has been collected to the database from October 2013 to October 2014. Each Region has been responsible for the reporting in process and ongoing telemedicine projects. The regions and the municipalities have had the opportunity to enter the data directly into the database or sending the data to MedCom who then has completed the data entry process. MedCom has verified the data entered by the regions or the municipalities and the regions have verified the data entered by MedCom. For data viewing, the database has a query form interface as well as a graphic interface based on Google Maps. The complete content of the database can be downloaded in an excel format file.

Data Analysis

Parts of the data exist in the database as free text; however, a significant amount of the data is categorized and entered as numerical or dichotomous data. The analyses in this paper are limited to frequency counts of the central structured parameters and a few cross tabulations.

Results

372 telemedicine initiatives are included in the database. The majority of 204 initiatives are still in a project state, which means that development is still going on and they are financed by temporary funding. 157 are run in a daily operation modus, and 11 initiatives are in a stage of dissemination to regional or national coverage.

Disease Area

The initiatives cover a wide area of diseases. Figure 1 shows the distribution of the disease areas covered by the 372 initiatives. As some projects cover more than one area, this explains why the sum is exceeding the total number of initiatives.

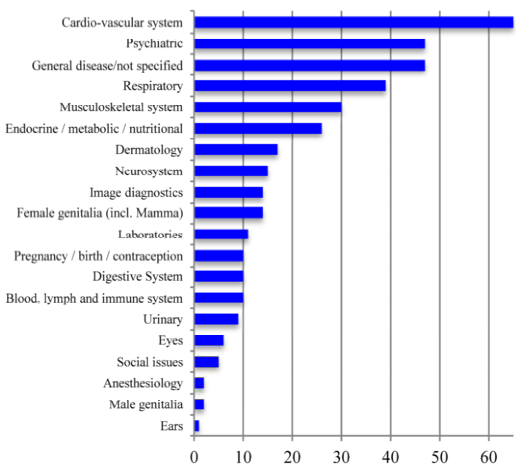


Figure 1 - Disease areas covered by the telemedicine initiatives in Denmark (n=372)

Cardiovascular diseases are the dominant area, closely followed by psychiatric diseases. The general disease area covers mainly general health surveillance in specific population groups, but can also include e.g., pain in geriatric areas, oncology in pediatric areas or ulcer assessment. Respiratory diseases are quite common for telemedicine initiatives, particularly COPD. Telemedicine systems for musculoskeletal diseases are predominantly directed at rehabilitation activities. Endocrine, metabolic and nutritional initiatives are various systems for diabetes patients. Denmark has a large program in telemedicine for ulcer surveilance where homenurses transmit images taken by a mobile phone or smartphone to the hospital based dermatologist or other health professionals.

Locality of Service

The telemedicine activities take place at many different localities, but the absolute predominant locality is in the hospital. 55% of all the initiatives include activities at hospital units. 31% have activities at municipal institutions (nursing homes, rehabilitation centres, social institutions etc.). 30% include activities at patients' or citizens' home. 6% take place at outpatient clinics, and 4% take place at GP' offices or local psychiatric units.

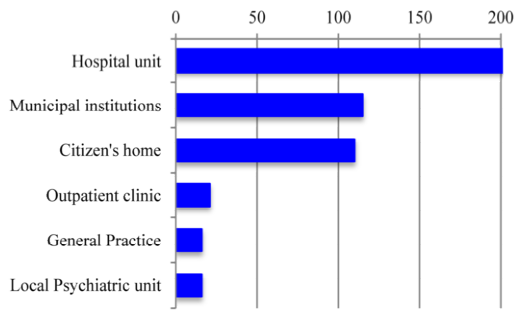


Figure 2 - Locality of service for telemedicine initiatives.

Hardware

Figure 3 shows what hardware is used in the telemedice initiatives.

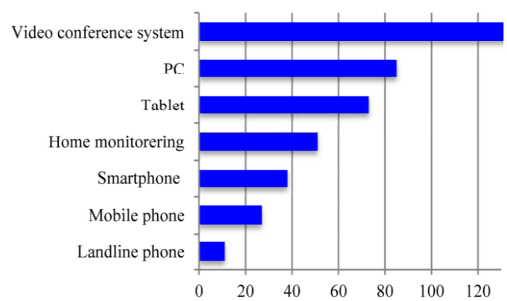


Figure 3 - Hardware used in the telemedicine initiatives

Video conference systems are used widely as they are included in many different tele services. Most of the cases as traditional video conferencing between various professional groups located in different hospitals, but also when teleinterpretation is necessary in the treatment of patients who do not speak Danish or English. Video cenferencing technologies are also applied in rehabilitation projects for instructing patients in their home, and to communicate with COPD patients.

Personal computers, tablets, and smartphones are used mainly in hospitals and municipal institutions by health professionals, whereas homemonitoring equipment is used in patients' home. Mobile phones and landline phones are used very little compared to the other more advanced technologies.

Integration to Other Health Care Data Systems

Few of the 372 telemedicine projects integrate to other health care data systems, as shown in Figure 4. 15 projects integrate to Electronic Health Record data in hospitals, and eight projects integrate to a GP system of which five also integrate to a hospital EHR. Seven projects integrate to homecare records in the municipalities, and six to the national health portal "Sundhed.dk".

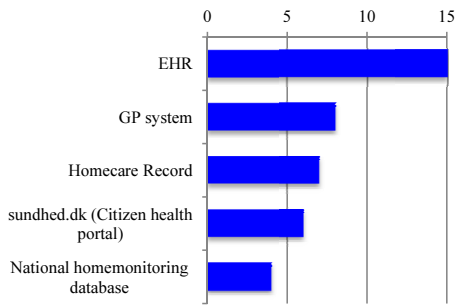


Figure 4 - Integration to other health care data systems

Discussion

It has been surprising how many projects have been reported to the database. Obviously the relatively broad definition of telemedicine from WHO can be interpreted to include a wide range of projects. For instance the high number of projects that use video conferencing will in a few years probably not be regarded as telemedicine projects, but just a normal part of professional communication as it is in many other sectors. The broad interpretations of telemedicine recur in other official documents in Denmark as well as in documents from private interest groups, not to mention the debate in general. From this perspective, the study cannot contribute with a precise presentation of the most important key aspects of a telemedicine solution – the spectrum remains very broad in the general discourse, as pointed out by Greenalgh et al [8].

The national strategy from 2012 [5] specifically mentions the following 5 initiatives as focus areas:

Clinically integrated home monitoring

Home monitoring for COPD patients

Demonstration and dissemination of telepsychiatry (Teleconferencing between health professionals)

Demonstration of Internet psychiatry (Depressions)

Dissemination of telemedical assessment of ulcers

The initiative on clinically integrated home monitoring was meant to include 2,000 patients in Denmark's largest cross-sector home monitoring project meant to run 2012-14. Patients with COPD, diabetes or inflammatory bowel disease as well as pregnant women with and without complications should be included in the project. The main objective was to provide experience of utilising a common technical solution for home monitoring of different groups of patients and to ensure coherent treatment procedures across the sectors in the health service. It was also the aim of the project to develop integration with the existing IT infrastructure, such as the EHR systems. Apart from a high number of projects concerned with cardiovascular diseases, which often is the monitoring of patients in their own home, it is difficult to see a significant concordance with the registrations in the database and the ambitious aims in the clinically integrated home monitoring initiative. As this is a cross section study, it is only possible to report on factual data from simple frequency counts and cross tabulations of the data available, hence it is not feasible to point to causal explanations of the observations.

Concerning the other initiatives there is a significant convergence between the strategy and the areas of diseases reported to the national telemedicine database.

The most predominant area of diseases is cardio-vascular disease. The projects in this category are often the monitoring of patients in their own home, but we also find many of the projects using information and communication technology to communicate between health professionals either individually or as a tele-video conference. The other main disease areas include the traditional chronic somatic diseases (Diabetes, COPD, Arthritis etc.) and the psychiatric area. The psychiatric area has two different categories of systems. One is termed telepsychiatry, which is video conferencing between psychiatric health care professionals to provide better courses of treatment for adult psychiatric patients. The other category, Internet psychiatry, is psychiatric treatment through an online IT program that is supposed to help patients suffering from depression. It is estimated that approximately 250,000 Danes suffer from depression, 30 percent of whom do not receive any treatment. It is estimated that four times more patients can be treated as in conventional treatment. The majority of the projects that indicate they cover general disease is mainly because they do not specify the area more precisely or because they cover more than one disease area, and therefore feel they are not covered by the list of disease areas.

The predominant localization for telemedicine projects registered in the database is professional institutions e.g. hospital units, municipal institutions, outpatient clinics, general practices, or psychiatric clinics. Activities in the patients' home are only stated by 106 of the 372 projects. This could underline the significance of the use of telemedicine to health professional communication.

The Methodological Approach

The data has been open for registration of projects to those responsible for the telemedicine projects in the five regions and 99 municipalities in Denmark. It has also been possible to send the data to MedCom to have them assist with data entry. However, as the decision on which projects to register has been up to the regions and municipalities, it is possible that some of the most active authorities have registered projects that are marginal to telemedicine, and more closely related to wellfare technologies. One example is a project focusing on the use of robot vacuum cleaners in the home of chronic deseased elderly citizens. Other projects might have been overlooked because of limited manpower in the municipalities and regions.

There are also large variations in the categorizations of the single parameters. Some projects have been very active in reporting - all the different technologies involved and, all the disease areas that they might face in the project, whereas other projects have underreported, omitting important fields. The validity of the database will be improved by a systematic validation of all the projects. However, the current state of the database is assessed to be valid for a broad characterization of the situation for telemedicine dissimination in Denmark.

Conclusion

In relation to the national strategy for telemedicine in Denmark, the database gives an overview of the initiatives in the regions and municipalities. It shows that the elements of the strategy are well prioritized following the guidelines in the strategy. The very large number of projects could also indicate a lack of national coordination.

A national coordinated evaluation of all the initiatives could improve the outcome of the projects in relation to both health outcome and cost related to development, implementation, and deployment.

References

[1] European Commission, Joint Research Centre, Institute for Prospective Technological Studies European Hospital Survey: Benchmarking Deployment of eHealth Services (2012-2013) 2014.

[2] HIMSS Europe, Electronic Medical Record Adoption in Denmark. Prepared August 28, 2014. Available at: http://regioner.dk/~/media/Mediebibliotek_2011/SUNDHED/Sundheds-it/HIMMS_Denmark_2014.ashx accessed 21. December 2014.

[3] Sindya N. Bhanoo, Denmark Leads the Way in Digital Care, New York Times, January 11, 2010.

[4] Protti D, Johansen I. Widespread Adoption of Information Technology in Primary Care Physician Offices in Denmark: A Case Study. The Commonwealth Fund. 2010.

[5] Telemedicine - A key to health services of the future - National action plan for the dissemination of telemedicine – in brief. Copenhagen: The Danish Agency for Digitisation (Digitaliseringsstyrelsen), 2012.

[6] https://medcom.medware.dk/tm/kort accessed 21. December 2014

[7] WHO. A health telematics policy in support of WHO's Health-For-All-Strategy for global health development: report of the WHO group consultation on health telematics, 11-16. December, Geneva, 1997. Geneva, World Health Organization, 1998.

[8] Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2:e001574. doi:10.1136/ bmjopen-2012-001574

Address for correspondence

Lead author contact: CN@plan.aau.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-501

Clinical Informatics Board Specialty Certification for Physicians: A Global View

Adi V. Gundlapalli[a, b], Aditya V. Gundlapalli[c], William W. Greaves[d], Denece Kesler[e], Peter Murray[f], Charles Safran[g], Christoph U. Lehmann[h]

[a]*University of Utah School of Medicine and* [b]*VA Salt Lake City Health Care System, Salt Lake City, UT, USA*
[c]*Utah County Academy of Sciences, Orem, UT, USA,* [d]*American Board of Preventive Medicine, Chicago, IL, USA,*
[e]*University of New Mexico School of Medicine, Albuquerque, NM,USA,* [f]*International Medical Informatics Association, Geneva, CH,*
[g]*Harvard Medical School, Boston, MA, USA,* [h]*Vanderbilt University, Nashville, TN, USA*

Abstract

Clinical informatics workforce development is a high priority for medicine. Professional board certification for physicians is an important tool to demonstrating excellence. The recent recognition of clinical informatics as a subspecialty board in the U.S. has generated interest and excitement among the U.S. informatics community. To determine the extent of similar programs in countries around the world, we performed literature searches with relevant keywords and internet searches of websites of informatics societies around the world for mentions or descriptions of certifications and reviewed publicly available sources. The U.S. certification was prominent in the recent published literature. Germany and Belgium have long-standing certifications with South Korea and Sri Lanka considering similar programs. This is the first global view of clinical informatics board certification for physicians. Training and certification for non-physician informatics professionals in allied areas are widespread. Official recognition and certification for physicians and all informatics professionals represents a key component of capacity building and a means of addressing the shortage of a skilled informatics workforce. Wider adoption of certification programs may further attracting talent and accelerate growth of the field.

Keywords:

Clinical informatics, board certification, physicians, informatics workforce.

Introduction

There is a global need for sustaining and growing the informatics workforce. A key aspect of fostering interest and attracting talented candidates is to provide opportunities for training and career advancement in the field. Voluntary 'certification' of professionals is a highly-visible quality indicator and a tool to improve physician recognition. Professionals from diverse backgrounds such as nursing, pharmacy, clinical medicine, and computer science work in the field of clinical or health informatics. While there are many opportunities for certifications in allied fields such as information technology and information systems, there are fewer opportunities in the field of informatics. In the U.S., nursing informaticians have had a certification program [1, 2] and there are efforts underway to establish an interprofessional certification [3]. While physicians are eligible to apply for certification pursued by informatics professionals, there are limited opportunities exclusively designed for physicians.

Specialty training and "certification" by a professional board after graduating from medical school represents a path for physicians to demonstrate expertise and dedication in a specific area of clinical medicine. This common route for physicians practicing in the US exists in comparable pathways in Africa, Asia, Americas, Australia, and Europe. Well-recognized board certification pathways exist in most countries for clinical specialties such as internal medicine, family practice, surgery, pathology, or radiology. Many countries offer subspecialties in medical and surgical fields. These pathways are considered important and in the US have become essential for practicing clinical medicine, hospital staff privileges, faculty appointments in schools of medicine and to establish appropriate credentials and qualifications with healthcare purchasers [4].

Clinical informatics was recognized in 2011 as a medical subspecialty in the U.S. for physicians. The modern seeds were sown in 2005 during a town hall meeting conducted by the American Medical Informatics Association (AMIA) [5]. Subsequent development of the core content and fellowship requirements by leading informatics professionals [6, 7] formed the foundation of the recognition of clinical informatics as a distinct subspecialty [8, 9]. The subspecialty certification is co-sponsored by the American Board of Preventive Medicine and the American Board of Pathology. Until 2017, it is available to any physician who possesses a license to practice and an unexpired board certification in any other specialty from the American Board of Medical Specialties and can demonstrate more than 25% clinical informatics efforts for three of the last five years.

One of the motivations for the U.S. subspecialty certification was the recognition that informatics is now considered an essential component to the practice, education, and research aspects of all medical specialties and subspecialties [10, 11]. It is anticipated that in the short-term, the ability to show competency and expertise in this new field will act as a catalyst for the training and recruitment of experts to advance clinical informatics in hospitals and practices. In the long term, certification should allow for uniformity and standardization in training for physicians and prepare expert clinical informaticians. It is reasonable to assume that the desire and need to have qualified physician informatics specialists to fill positions such as chief medical/health informatics officers, directors of clinical informatics, and physician leads of EHR implementations will increase in the future.

With the excitement generated by the U.S. board certification [8], we sought to review existing data on the status of clinical informatics as a specialty or subspecialty for physicians outside the US. The hypothesis was that countries with well-established informatics infrastructure will have similar certification programs for physicians.

Methods

Literature Search

We used combinations of the following keywords for literature searches: "clinical informatics", "health informatics", "biomedical informatics", "specialty" or "subspecialty", "board certification", "physicians", "doctors" to search PubMed, Medline, Scopus, Web of Science, World Cat, CINAHL and Google Scholar. These databases were accessed through the University of Utah intranet.

Internet Search of Informatics Society Web Sites

A listing of member societies of the International Medical Informatics Association (IMIA) was reviewed as of November 30, 2014 from the IMIA website [12]. Brief descriptions of the member societies as listed on the IMIA page were first reviewed. Subsequently, individual society websites were accessed and contents reviewed for evidence of clinical informatics certifications or qualifications with an emphasis on physicians. Websites in languages other than English were reviewed using the automatic translation feature of either Google Chrome or Bing through Internet Explorer. Complying with rules for good scientific practice, all webpage screenshots or pdf versions used for this publication were archived.

Informal Discussions with Informaticians at International Meetings

The authors discussed the topic of board certification in clinical informatics for physicians with informatics professionals at the 2014 Asia Pacific Association of Medical Informatics (APAMI) held in New Delhi, India in early November 2014 and at the Annual Symposium of the American Medical Informatics Association (AMIA) in Washington, DC in mid-November 2014.

Results

Results of Literature Search

In reviewing the literature, the concept of an informatics subspecialty is not entirely new and was raised as early as 1985 and 1993 [13, 14]. The forward looking vision of Kunstaetter in 1985 is impressive [14]: *"The medical profession has to become directly involved by establishing and supporting medical informatics as a new specialty. To do otherwise would be equivalent to leaving the practice of radiology to physicists or medical therapeutics to the pharmaceutical industry."*

The search of databases yielded few relevant results related to clinical informatics board certification for physicians; as shown in Table 1, most of the recent papers were related to the US experience [5-11, 15]. The papers trace the history of the US board certification from concept to setting requirements to administration of the examination. The US certification generated considerable excitement in the US physician informatics community. The next step in this process is the establishment and sustenance of accredited fellowship programs for clinical informatics that will train the next generation of clinical physician informaticians.

During the literature search for clinical informatics certification, we noted separate pathology informatics training for pathologists in the US [16-18]. It is important to note that pathologists are eligible to apply for the US subspecialty board certification in clinical informatics.

Roger France *et al* review the certification process in existence in Belgium for physicians since an official ministerial decree was passed in 2001 by the Belgian parliament [19]. The criteria for being designated a "*Physician Specialist in Health Data Management"* in Belgium include being a licensed physician and requirements for formal coursework and practical training (each one year) and presentation of an original dissertation. There are no published articles regarding the success and challenges of this program nor an estimate of the number of physicians who hold this certification.

Two articles describe a "supplement medical informatics" qualification for German physicians first approved in 1991 [20, 21]. This qualification requires 1.5 years of formal coursework and practical experience that has to be certified by the physicians' institutional leadership. For this certification, we were unable to find published data on the number of physicians who hold this certification.

South Korea has established a formal training program in biomedical informatics for physicians [22]. The 18-month program leads to Certified Physicians in BioMedical Informatics (CPBMI), certified by the Korean Society of Medical Informatics (KOSMI). Kim anticipates that the next step would be to establish a board certification in biomedical informatics similar to the US program. There are numerous academic training programs (degree and non-degree granting) that are open to physicians with recommendations for standards and accreditation [23-25]. Several countries offer certification and recognition processes for informatics professionals including physicians [26, 27].

Table 1. Results of database searches for literature pertaining to clinical informatics board certification specifically for physicians (see text for combinations of keywords used)

Databases	Relevant articles from manual review of main search and 'related searches' (and top 100 results from Google Scholar)
PubMed, Medline, Scopus, Web of Science, World Cat, CINAHL and Google Scholar	Applicable specifically to physician certification: US related [5-11, 15]; World: [19-22]

Results of Internet Search of Informatics Society Web Sites

A review of the brief descriptions of the member societies charters on the IMIA website revealed training and career development of informatics professionals as an often cited goal with no specific mentions of certification specifically for physicians.

We reviewed the websites of 58 member societies of informatics listed on the IMIA website and 5 regional member associations. Countries listed as "Corresponding members" did not have websites listed. As shown in Table 2, most were amenable to review by virtue of being in English or translated using either Google Chrome or Bing. Information on certification specifically for physicians was noted on the US site (American Medical Informatics Association, AMIA). Information on the German site matched the literature search [20, 21], as did the link to the certification for physicians on the South Korean site [22]. There was no mention of the certification available in Belgium on their website. All other accessible websites had no mention of certification specifically for physicians.

Many countries offer certification for all informatics professionals and physicians would likely be eligible for those training and certification processes. Examples include the UK, Australia, and Canada.

The IMIA website and those of the Asia Pacific Association of Medical Informatics, European Federation for Medical Informatics, and Pan African Health Informatics Association yielded no information on certification specifically for physicians. The Regional Federation of Health Informatics for Latin America and the Caribbean website did not load and the Middle East Association for Health Informatics had no website listed.

Results of Informal Discussions with Informaticians

A meeting with the President of the Health Informatics Society of Sri Lanka resulted in our being alerted to the existence of the Specialty Board in Biomedical Informatics in that country at the Postgraduate Institute of Medicine at the University of Colombo and the possibility of a board certification for physicians in informatics in the near future (personal communication, Prof. Vajira H. W. Dissanayake). In performing internet searches on this topic, the Sri Lankan society on the IMIA webpage briefly mentions a master's course in biomedical informatics offered by the Specialty Board in Biomedical Informatics that is specifically offered for medical doctors and dentists in Sri Lanka [28].

Informal discussions with members of the editorial boards of the Applied Clinical Informatics journal and the International Journal of Medical Informatics at the 2014 AMIA Annual Symposium indicated no board certification pathways in clinical informatics for physicians in Brazil or Australia.

Discussion

Over the years, clinical informatics has had a significant impact on the practice of medicine. Demand for increasing quality and efficiency, while decreasing costs and errors, requires an informed and well-trained workforce in clinical informatics. As in any field, we face challenges in recruiting and retaining talented professionals to clinical informatics.

Table 2. Review of International Association of Medical Informatics member society websites for information on clinical informatics-related board certification opportunities and pathways specifically for physicians

Continent	Results/Comments
Africa	Cameroon, Kenya, Mali, Nigeria, South Africa: No information on clinical informatics board certification for physicians Burundi, Côte d'Ivoire, Ghana, Malawi, Togo: No website listed
Americas	Canada, Chile, Cuba, Uruguay: No information on clinical informatics board certification for physicians Argentina, Mexico, Peru: Website failed to load Brazil: Page could not be translated Colombia: No website listed ***USA: Information on US board certification for physicians***
Asia	China, Hong Kong, India, Japan, Pakistan, Philippines, Singapore, Taiwan, Thailand: No information on clinical informatics board certification for physicians Israel: No website listed Sri Lanka: Website failed to load Iran, Saudi Arabia: Website under construction ***South Korea: Link to certification for physicians in Biomedical Informatics***
Australia	Australia, New Zealand: No information on clinical informatics board certification for physicians
Europe	Belgium, Croatia, Czech Republic, Great Britain, Greece, Hungary, Finland, France, Ireland, Norway, Romania, Slovenia, Spain, Switzerland, Sweden, The Netherlands, Turkey, Ukraine: No information on clinical informatics board certification for physicians Austria: Page could not be translated Bosnia and Herzegovina, Italy: Page failed to load ***Germany: Medical informatics certification specifically for physicians***

A formal certification process with subsequent tangible benefits such as official acknowledgement and recognition of excellence, qualification for a named position of authority and possibly monetary benefits would go a long way in attracting and retaining professionals to this field.

Physicians are an integral part of the clinical informatics team that consists of dedicated professionals from various disciplines. While physicians are likely satisfied to be recognized for their knowledge, skills, and experience in informatics, it would also be important to recognize those that have achieved

official board certification in their chosen field. In this context, the US clinical informatics subspecialty board certification fulfills a long awaited aspirational need and has generated much excitement and discussion [29, 30].

With the news of the recent U.S. certification, we set out to find other similar programs in countries worldwide. As with the U.S., short- and long-term training and degree granting programs exist for informatics in many countries and these are open to physicians. It was more challenging to determine if there are programs that are reserved and specifically designed for physicians.

Our hypothesis that countries with well-established informatics infrastructure will have similar certification programs for physicians was not validated. It was interesting to note that the US is the latest to join a very short list of countries such as Germany and Belgium that have had programs equivalent or similar to US board certifications for physicians in the field of clinical informatics for many years. With South Korea and Sri Lanka actively considering similar programs, there appears to be an opportunity for other countries to consider and organize their training to offer similar recognition. The motivations and tangible returns will likely vary for different countries as will the infrastructure, logistics, social and political will to establish such programs.

We acknowledge several limitations of our study. Key word searches of online literature databases may be incomplete based on filtering for English language articles and choice of keywords. Our search would have missed the non-English literature. The automatic translation of non-English language websites was not independently verified by those familiar with the language and thus we may have missed references to board certification pathways. There were some sites that were not amenable to translation from their native language.

As this topic generates more interest among the international informatics community, there may also be opportunities to formally engage the IMIA member societies in dialog regarding certification opportunities and pathways for physicians. This could be conducted via email, online, or in-person surveys at international informatics meetings. IMIA might even serve as an authority on certification for member societies.

We encourage and request individuals with knowledge and experience with training and certification programs exclusively tailored for physicians in different countries to email us with details. We also encourage stewards of national informatics societies to email us with details of clinical informatics related certifications and qualifications for physicians in their countries. It would be important to have an exhaustive and as-complete-as-possible inventory of such programs so that best practices, motivations and lessons learned could be shared among informatics professionals.

Acknowledgements

We wish to acknowledge our international informatics colleagues who freely discussed this topic with the authors at various informatics meetings. We wish to appreciate the efforts of prominent informatics leaders from AMIA who worked tirelessly to pursue the US board certification and the American Board of Preventive Medicine for taking the lead to make the certification a reality. We thank Marjorie Carter (VA Salt Lake City Health Care System) for editorial and formatting assistance. There was no specific grant funding for this project.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the US Department of Veterans Affairs, the United States government, the American Board of Preventive Medicine, the International Medical Informatics Association, or our respective universities.

References

[1] V. K. Saba, D. J. Skiba, and C. Bickford, "Competencies and credentialing: nursing informatics," *Stud Health Technol Inform,* vol. 109, pp. 75-89, 2004.

[2] J. G. Ozbolt and V. K. Saba, "A brief history of nursing informatics in the United States of America," *Nurs Outlook,* vol. 56, pp. 199-205 e2, Sep-Oct 2008.

[3] American Medical Informatics Association. (2014). *Advanced Interprofessional Informatics Certification (AIIC) Program*. Available: http://www.amia.org/advanced-interprofessional-informatics-certification

[4] American Board of Medical Specialties. (2014). *Board Certification: A Trusted Credential*. Available: http://www.abms.org/board-certification/a-trusted-credential/

[5] D. E. Detmer, J. R. Lumpkin, and J. J. Williamson, "Defining the medical subspecialty of clinical informatics," *J Am Med Inform Assoc,* vol. 16, pp. 167-8, Mar-Apr 2009.

[6] R. M. Gardner, J. M. Overhage, E. B. Steen, B. S. Munger, J. H. Holmes, J. J. Williamson*, et al.*, "Core content for the subspecialty of clinical informatics," *J Am Med Inform Assoc,* vol. 16, pp. 153-7, Mar-Apr 2009.

[7] C. Safran, M. M. Shabot, B. S. Munger, J. H. Holmes, E. B. Steen, J. R. Lumpkin*, et al.*, "Program requirements for fellowship education in the subspecialty of clinical informatics," *J Am Med Inform Assoc,* vol. 16, pp. 158-66, Mar-Apr 2009.

[8] D. E. Detmer and E. H. Shortliffe, "Clinical informatics: prospects for a new medical subspecialty," *JAMA,* vol. 311, pp. 2067-8, May 28 2014.

[9] E. H. Shortliffe, "President's column: subspecialty certification in clinical informatics," *J Am Med Inform Assoc,* vol. 18, pp. 890-1, Nov-Dec 2011.

[10] D. E. Detmer, B. S. Munger, and C. U. Lehmann, "Clinical informatics board certification: history, current status, and predicted impact on the clinical informatics workforce," *Appl Clin Inform,* vol. 1, pp. 11-8, 2010.

[11] C. U. Lehmann, V. Shorte, and A. V. Gundlapalli, "Clinical informatics sub-specialty board certification," *Pediatr Rev,* vol. 34, pp. 525-30, Nov 2013.

[12] International Medical Informatics Association. (2014). *International Medical Informatics Association: Member Societies*. Available: http://www.imia-medinfo.org/new2/member_societies

[13] R. D. Aller, "Clinical informatics as a medical subspecialty," *Healthc Inf Manage,* vol. 7, pp. 11-6, Fall 1993.

[14] R. Kunstaetter, "The need for a new specialty: medical informatics," *Can Med Assoc J,* vol. 133, pp. 236-8, Aug 1 1985.

[15] P. Graham-Jones, S. H. Jain, C. P. Friedman, L. Marcotte, and D. Blumenthal, "The need to incorporate health information technology into physicians' education and professional development," *Health Aff (Millwood),* vol. 31, pp. 481-7, Mar 2012.

[16] J. H. Sinard, S. Z. Powell, and D. S. Karcher, "Pathology training in informatics: evolving to meet a growing need," *Arch Pathol Lab Med,* vol. 138, pp. 505-11, Apr 2014.

[17] B. P. Levy, D. S. McClintock, R. E. Lee, W. J. Lane, V. E. Klepeis, J. M. Baron*, et al.*, "Different tracks for pathology informatics fellowship training: Experiences of and input from trainees in a large multisite fellowship program," *J Pathol Inform,* vol. 3, p. 30, 2012

[18] A. B. Carter, "Board certification for pathologists in clinical informatics: Are you a lumper or a splitter?," *J Pathol Inform,* vol. 3, p. 12, 2012.

[19] F. H. Roger France, C. Beguin, C. Melot, and P. Gillet, "Board certified physicians in health informatics a European precedent for professional recognition," *Yearb Med Inform,* pp. 116-20, 2010.

[20] R. Haux, J. Dudeck, W. Gaus, F. J. Leven, H. Kunath, J. Michaelis*, et al.*, "Recommendations of the German Association for Medical Informatics, Biometry and Epidemiology for education and training in medical informatics," *Methods Inf Med,* vol. 31, pp. 60-70, Feb 1992.

[21] J. Michaelis, "Professional qualification of German physicians in medical informatics," *Methods Inf Med,* vol. 33, pp. 312-4, Jun 1994.

[22] J. H. Kim, "Certification for physicians in biomedical informatics," *Healthc Inform Res,* vol. 19, pp. 1-2, Mar 2013.

[23] E. Manifava, A. Kolokathi, and J. Mantas, "Exploring the biomedical and health informatics educational programs in europe," *Stud Health Technol Inform,* vol. 202, pp. 67-70, 2014.

[24] A. Hasman and J. Mantas, "IMIA Accreditation of Health Informatics Programs," *Healthc Inform Res,* vol. 19, pp. 154-61, Sep 2013.

[25] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh*, et al.*, "Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision," *Methods Inf Med,* vol. 49, pp. 105-120, Jan 7 2010.

[26] H. F. Marin, E. Massad, E. P. Marques, R. S. Azevedo, and L. Ohno-Machado, "Training health informatics professionals in Brazil: rationale for the development of a new certificate program," *AMIA Annu Symp Proc,* p. 1042, 2005.

[27] P. McCullagh, G. McAllister, P. Hanna, D. Finlay, and P. Comac, "Professional development of health informatics in Northern Ireland," *Stud Health Technol Inform,* vol. 169, pp. 218-22, 2011.

[28] Specialty Board in Biomedical Informatics. (2013). *Master of Science in Biomedical Informatics*. Available: http://pgim.cmb.ac.lk/wp-content/uploads/2014/09/BMI-prospectus-29.11.2013.pdf

[29] E. Digitale. (2014). *5 Questions: Longhurst on clinical informatics gaining recognition as medical sub-specialty*. Available: http://med.stanford.edu/news/all-news/2014/01/5-questions-longhurst-on-clinical-informatics-gaining-recognition-as-medical-sub-specialty.html

[30] W. Hersh. (2012). *Informatics Professor Blogspot: Challenges for Building Capacity of the Clinical Informatics Subspecialty* Available: http://informaticsprofessor.blogspot.com/2012/09/challenges-for-building-capacity-of.html

Address for correspondence

Adi V. Gundlapalli, MD, PhD, MS
Associate Professor
Departments of Internal Medicine and Biomedical Informatics
University of Utah School of Medicine
Chief Health Informatics Officer
VA Salt Lake City Health Care System
Adi.Gundlapalli@hsc.utah.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-506

A Pilot Study of Computer-Based Simulation Training for Enhancing Family Medicine Residents' Competence in Computerized Settings

Aviv Shachak[a,b], Sharon Domb[c,d], Elizabeth Borycki[e], Nancy Fong[b], Alison Skyrme[b], Andre Kushniruk[e], Shmuel Reis[f], Amitai Ziv[g]

[a] *Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada*
[b] *Faculty of Information, University of Toronto, Toronto, ON, Canada*
[c] *Sunnybrook Health Sciences Centre, Toronto, ON,Canada*
[d] *Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada*
[e] *School of Health Information Sciences, University of Victoria, Victoria, BC, Canada*
[f] *Faculty of Medicine in the Galilee, Bar-Ilan University, Safed, Israel*
[g] *MSR- Israel Center for Medical Simulation, Sheba Medical Center, Tel-Hashomer, Israel*

Abstract

We previously developed a prototype computer-based simulation to teach residents how to integrate better EMR use in the patient-physician interaction. To evaluate the prototype, we conducted usability tests with three non-clinician students, followed by a pilot study with 16 family medicine residents. The pilot study included pre- and post-test surveys of competencies and attitudes related to using the EMR in the consultation and the acceptability of the simulation, as well as 'think aloud' observations. After using the simulation prototypes, the mean scores for competencies and attitudes improved from 14.88/20 to 15.63/20 and from 22.25/30 to 23.13/30, respectively; however, only the difference for competencies was significant (paired t-test; t=-2.535, p=0.023). Mean scores for perceived usefulness and ease of use of the simulation were good (3.81 and 4.10 on a 5-point scale, respectively). Issues identified in usability testing include confusing interaction with some features, preferences for a more interactive representation of the EMR, and more options for shared decision making. In conclusion, computer-based simulation may be an effective and acceptable tool for teaching residents how to better use EMRs in clinical encounters.

Keywords:

Education; Electronic Medical Record; Physician-Patient Relationship; Simulation-Based Training.

Introduction

The potential and actual benefits of health information technology in general, and electronic medical records (EMRs) in particular, have been widely discussed. However, some concerns have been raised over unintended consequences and especially the impact of EMRs on patient-clinician communication [1]. Communication skills are central to patient-centered care and have been associated with patient satisfaction, conflict resolution, adherence to treatment, and a myriad of health outcomes [2]. Research has shown that the use of EMRs affects the patient consultation in both positive and negative ways. On the positive side, the use of EMR improves the exchange of medical information between physicians and patients. However, it often interferes with maintaining eye contact, establishing rapport, and psychological and emotional communication. Furthermore, physicians rarely utilized resources within the EMR, and the computer in general, for patient education [3].

Currently, the training provided to clinicians focuses mostly on technical aspects (e.g. how to document, how to prescribe a medication electronically) of using the EMR and not on how to best integrate it into the patient consultation; however, the need to "go beyond the nuts and bolts (operator skills) of using new technologies" [4] has been recognized. In previous studies, we identified some of the cognitive issues underlying the impact of EMRs on patient-physician communication, as well as strategies and best practices employed by physicians in order to overcome the negative and maximize the positive influences of the EMR on the consultation. Based on these findings, we developed and tested a simulation-based training intervention with standardized patients (actors) aimed at enhancing family medicine residents' competence in computerized settings [5]. However, widespread implementation of this simulation is compromised by its cost and scalability. Therefore, we developed a prototype computer-based simulation, named EMR-sim, which can be widely distributed and implemented [6]. The purpose of this study was to pilot test the prototype computer-based simulation in an attempt to: 1) identify usability and design issues, 2) examine its impact on family medicine residents' self-reported competencies and attitudes related to using the EMR in the consultation, and 3) assess the acceptability of the simulation to its intended audience (family medicine residents).

Methods

The Computer Based Simulation: EMR-sim

We described the development of the computer-based simulation—EMR-sim—elsewhere [see 6]. The current version of our EMR-sim is a Flash -based application that runs in a browser. To keep the cost minimal, it was developed using Adobe Captivate 7, which does not require extensive knowledge in programming and employs basic graphics. The application presents the user with screen captures from the EMR that can be enlarged by hovering a mouse over a magnifying glass image, dialogue texts, and decision buttons (Figure 1). The scenarios captured such issues as dealing with privacy and safety concerns related to documenting information in the wrong patient's chart [7], communicating with a triadic patient who may be distracted by the computer

[8], and using the EMR for patient education (e.g. by visualizing trends in lab results and use of risk calculators). After completing each scenario, the trainee is presented with feedback, specifically tailored to his or her choices, and references to support the feedback.

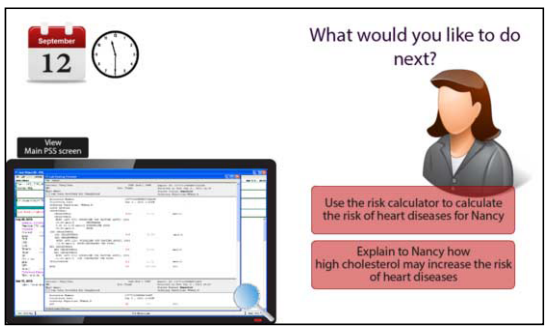


Figure 1- Screen Capture from EMR-sim

Phase 1: Pre-Pilot Usability Testing

A pre-pilot usability study was conducted as part of a course on usability evaluation at the Faculty of Information (iSchool) of the University of Toronto. Two of the authors (NF, ASk) and a third student conducted the pre-pilot under the guidance of the course instructor. A convenience sample of three Master of Information students with no clinical experience served as participants. They were assessed individually in separate 40-45 minutes sessions. Each session began with an introduction to the study and EMR-sim, followed by 'think aloud' observation [9] in which participants interacted with the simulation prototypes while verbalizing their actions, thoughts, and feelings. One of the researchers facilitated the observation using prompts such as "What are you trying to do now?," "What did you expect to happen (when you…)?," or "How did you feel about that?." A second researcher took notes.

Phase 2: Pilot Study

Participants' Recruitment and Data Collection

We sent email invitations to participate in the study to 49 Family Medicine residents from two teaching hospitals in Toronto (Sunnybrook Health Sciences Centre and Women's College Hospital), with a reminder email two weeks later. 16 residents consented (response rate: 32.7%) and were enrolled in 2 similar one-hour evaluation sessions (one with 11 and the other with 5 participants) during the lunch break (with lunch provided). At the end of the evaluation session participants received a $50 gift card honorarium.

At the beginning of each session, the purpose of the study was explained and the informed consent form was reviewed with the participants. After signing this form, participants filled in a pre-test questionnaire comprised of 5 items for self-reported competencies (measured on a 4-point scale ranging from low to excellent), and 6 attitudes items (measured on a 5-point Likert scale of agreement) related to using the EMR in the consultation. Demographic characteristics of participants (gender, age group, and year of residency) were also collected on the pre-test questionnaire.

After completing the first survey, participants interacted with the simulation prototype for two scenarios and then filled in a post-test questionnaire. The post-test questionnaire included the same measures for competencies and attitudes, plus additional items to assess the acceptability of the simulation and overall rating of the session. It also included a space for free text comments on the session.

Items measuring competencies, attitudes, and overall evaluation of the session (SEVAL) were taken from our previous study with standardized patients [5], with some modifications. Acceptability of the simulation was measured using the Technology Acceptance Model (TAM) [10]. TAM is a widely used model, which consists of 3 constructs: Perceived Usefulness (PU), Perceived Ease of Use (PEOU) and intention to use the system (INT). Although other instruments such as the Unified Theory of Acceptance and Use of Technology (UTAUT) typically explain greater portion of the variance in usage intention [11], TAM is a useful tool for practical purposes such as assessing the acceptability of an information system because of its parsimony. Our survey instrument included 6 items for perceived usefulness, 3 for perceived ease of use, and 4 items for usage intention. During the sessions, we observed 6 of the participants (3 in each session, 5 females and 1 male) as they interacted with the simulation and asked them to 'think aloud' as described before. The observers took notes and participants were audio recorded.

Prior to beginning the research, the study was approved by the research ethics boards of Sunnybrook Health Sciences Centre and the University of Toronto, and all participants gave written informed consent.

Quantitative Data Analysis

Reliability (Cronbach's alpha) of the measures for self-reported competencies and attitudes related to using the EMR in the consultation were tested using the pre-test data. Perceived usefulness, perceived ease of use, intention to use the simulation and participant's evaluation of the session were only measured on the post-test, and their reliability was tested using the post-test data. Except for attitudes towards using the EMR in the consultation that was low, scale reliability of all other measures was acceptable or good (Table 1).

Table 1 – Scale reliabilities of the measure instrument

Measure	Reliability (Cronbach's alpha)
Patient-physician-EMR competencies	0.75
Attitudes towards using the EMR in the consultation	0.28
Perceived usefulness of the simulation	0.92
Perceived ease of use of the simulation	0.76
Intention to use the simulation	0.80
Overall evaluation of the session	0.75

Data were analyzed using IBM SPSS Statistics 22. Combined scores for competencies and attitudes were calculated by summarizing the scores of all individual items and average scores were calculated for perceived usefulness, perceived ease of use, intention, and evaluation of the session. After testing for normality of the data distribution, we used paired-samples t-tests to compare competencies and attitudes scores before and after using the simulation prototypes. Effect sizes were calculated using G*Power 3.0.10 software. Inter-rater agreements on perceived usefulness, perceived ease of use, intention, and evaluation of the session were calculated by entering the data into MS-Excel and using the rWG(j) formula [12].

Qualitative Data Analysis

Qualitative data were analyzed as follows. First, researchers read through the free-text comments and notes from the 'think aloud' observations to familiarize themselves with the data. The three researchers who conducted 'think aloud' observations (AS, NF, and AS) met to review their notes for

clarification, compare the notes, and discuss preliminary themes that emerged. Then, free-text comments, notes, and audio recordings of the 'think aloud' sessions were entered into NVivo 8 qualitative data analysis software. The first author applied open coding to the data, adding new codes as they emerged. Codes were then grouped into categories and reorganized to develop the final coding scheme. The first author then used the scheme to recode all the data. To establish trustworthiness, a second researcher (EB) also coded a subset of the data (the free-text comments) using the same coding scheme. There was high agreement on coding (91% as calculated by NVivo) and moderate agreement on categorization (κ=0.56). The two researchers then discussed their interpretations, following which consensus on coding and themes had been reached.

Results

Pre-Pilot Usability Testing (Phase 1)

As participants were not clinicians, they gave little feedback on the content of the simulations and focused mainly on the navigation, clarity, and overall usability of the scenarios. Positive findings included reaction to emotions shown by the patient's avatar, particularly when tears were shed. All participants found this to be very effective.

The pace of the doctor-patient dialogue was one of the main issues identified by our participants. The unnaturally slow pace caused users to attempt to rush through the dialogue and subsequent navigational prompts and miss certain cues. Another key issue was a lack of clarity of system navigation. The original prototypes included a video play bar, which is added by Adobe Captivate by default. The majority of participants attempted to interact with this bar, instead of using the decision buttons, to progress through the scenario. A third issue was the 'mouse over' interaction with the magnifying glass that was confusing.

Based on these findings, we modified the pace of the dialogue and removed the video bar from the prototypes. The interaction with the magnifying glass is more difficult to change. We decided not to change it at this time but explain this mode of interaction at the beginning of the pilot sessions.

Characteristics of Study Participants

Descriptive statistics of the phase 2 pilot study participants are presented in Table 2. Of the 16 Family Medicine residents who participated in the study, the majority (n=14) were women. Except for one participant who was 40 or older, all participants were between ages 20 and 39—most of them (n=9) in their twenties. The majority of participants (n=10) were in their first year of residency (PGY1).

Table 2 – Characteristics of phase 2 study participants (N=16)

Participant characteristic	N
Gender	
Female	14 (81.2%)
Male	2 (12.5%)
Age group	
20-29	9 (56.3%)
30-39	6 (37.5%)
40-49	1 (6.3%)
Year of residency	
1st year	10 (62.5%)
2nd year	6 (37.5%)

Effect of Using the Simulation Prototypes on Skills and Attitudes

The combined score for self-reported competencies of integrating the EMR into the patient consultation improved from 14.88±2.63 before to 15.63±2.80 (M±SD; out of maximum 20 points possible) after using the simulation prototypes. This difference was statistically significant with a large effect size (Table 3). When the scores for individual items were compared, the scores for 4 of 5 items increased from pre- to post-simulation, but this improvement was only significant for 2 items: technical aspects of operating the EMR during the patient visit (paired t-test; t=-2.611, p=0.020) and using the computer for patient education (paired t-test; t=-2.611, p=0.020), which improved from 2.63±0.81 to 2.94±0.77 and from 2.75±0.86 to 3.06±0.77 (M±SD; on a 4-point scale), respectively. The score for one item (prevention of EMR-related errors) dropped from 3.31±0.60 to 3.06±0.68 (M±SD; on a 4-point scale), but this was not statistically significant (paired t-test; t=1.73, p=0.104).

The combined score for attitudes related to using the EMR during the patient visit also increased from 22.25±2.44 before to 23.13±2.16 (M±SD; out of maximum 30 points possible) after using the simulation prototypes. However, this difference was not significant (Table 3). The scores for 5 of 6 attitude items improved from pre- to post-simulation and decreased for one item ("use of the EMR enhances my performance"); however, none of these changes were statistically significant.

Table 3 – Self-reported competencies and attitudes related to the use of the EMR in the consultation.

	Pre-test		Post-test		t(df=15)	Sig.	Effect size (dz)
	M	SD	M	SD			
Competencies[a]	14.88	2.63	15.63	2.80	-2.535	0.023	0.63
Attitudes[b]	22.25	2.44	23.13	2.16	-1.754	0.100	0.44

[a] Combined score out of maximum 20 points possible.
[b] Combined score out of maximum 30 points possible.

Analysis of the free-text comments from the post-simulation questionnaire supports the positive impact of the computer-based simulation. In particular, respondents commented that the simulation increased their awareness of issues surrounding the use of EMR in the consultation *("Awareness of EMR pros and cons"* [p. 4]), and that it provided them with reassurance for their current practices *("Served to reinforce that my current practice is consistent with recommended/ acceptable practices"* [p. 7]).

Acceptability of the Simulation

The mean scores for acceptability of the simulation and overall evaluation of the simulation session, as measured on the post-simulation questionnaire, are presented in Table 4. Scores for perceived ease of use and perceived usefulness were good (4.10±0.73 and 3.81±0.74 on a 5-point scale, respectively). Intention to use the simulation and overall evaluation of the session were rated just above average. Inter-rater agreement on all scores was high (rWG(j)≥0.8)

Table 4 – Acceptability of the simulation and overall evaluation of the session.

	PU		PEOU		INT		SEVAL	
	M	SD	M	SD	M	SD	M	SD
	3.81	0.74	4.10	0.73	3.50	0.63	3.16	0.79
rWG(j)	0.90		0.82		0.89		0.80	

The qualitative data analysis provided additional support for the acceptability of the computer-based simulation. Participants commented that the scenarios were realistic, that the prototypes were easy to use and enjoyable *("Great, fun, simulation that is very realistic"* [p.7]), and that it would be effective for teaching Family Medicine residents how to better use the EMR in clinical encounters. However, they thought it would be most effective in earlier stages of the Family Medicine residency program: *"Use this to teach all residents! Would have been an excellent tool to have in the first month of residency. I would have definitely used this tool often when first learning how to use an EMR ..."* [p. 5]. Two participants, however, thought it would be difficult to teach these issues using a software application that cannot fully capture the nuances and range of patient behaviors: *"I find it is difficult to teach you how to balance using the computer during the actual encounter with patient interaction through an app, as the balancing act requires 'learning through doing' and is better done at clinic"* [p.12].

Usability and Scenario Design Issues

Several usability and scenario design issues emerged from the 'think aloud' observations. First, it was noted that, as opposed to the Master of Information students who focused mostly on the software itself, residents concentrated on the clinical aspects of the simulation; i.e., they spent substantial amount of time reviewing the information presented on the EMR and commented on the clinical decision options. Participants commented that they would like the EMR representation in the simulation to be more interactive—much more like the real system they were using (*"Creating an interface where we can actually use EMR during the simulation session"* [p. 8, in response to the question "what would improve the simulation?"]). They also wanted to have additional choices available—especially for shared decision making (e.g. "*I would ask her if she would like… it's hard for me to choose what I recommend"* [p.1,'think aloud' session]).

The pace of the simulation, which has been modified following the pre-pilot, did not seem to be an issue for most residents, except for two who commented on it on the post-simulation survey—one of whom was also observed in the 'think aloud' and seemed to be a very fast reader. Interaction with the magnifying glass was still confusing for some of the participants. Finally, through the 'think aloud' observation, we discovered a number of additional usability and scenario design issues that were not captured in the pre-pilot. For example, we noted that in a scenario discussing recent discovery of hypercholesterolemia, the EMR's screen capture indicated that the patient was already taking medications for this problem, which was confusing to participants. Other minor issues included a blank references screen when no references for the feedback were available, and one slide transition that was too fast.

Discussion

The effect of using EMRs on patient-clinician relationships has gained considerable research attention. However, educational interventions aimed at improving clinicians' competence in computerized settings are still rare. As part of our effort to bridge this gap, we have developed several interventions, of which EMR-sim is one [5, 6].

The results of this pilot study suggest that a computer-based simulation would be a useful tool for teaching Family Medicine residents how to better integrate the use of EMRs—and the computer in general—into the patient consultation. The combined scores for both self-reported competencies and attitudes related to using the EMR in the consultation improved from pre- to post-simulation (although this change was only significant for competencies), and so did most individual item scores. Our data suggest that the simulation is acceptable to users with good scores for perceived ease of use and usefulness. Free-text comments provided by participants support this analysis and suggest that the simulation may be especially useful in the early stages of the Family Medicine residency program. As a computer-based simulation is more scalable, compared to previous interventions such as simulation with standardized patients [5], we believe that it is worth further development and testing.

A pre-pilot usability testing with Master of Information students and a pilot test with Family Medicine residents proved to be a useful combination, as the students focused more on the interaction with the tool itself, whereas the residents examined mostly the medical information and decisions, and the display of the EMR component on the simulation prototype. This allowed us to identify some human-computer interaction issues, such as the pace of the conversation and interaction with the magnifying glass, as well as issues related to decision choices and interaction with the EMR component of the simulation prototype. In particular, residents pointed out that they would like to have more options for shared decision making. It has been suggested that the EMR can be effectively utilized for sharing understanding between patients and clinicians [13]. Additional scenarios that address this issue should be developed and tested. Second, residents preferred the simulated EMR to be more interactive—much like the real system they use. A potential way to achieve this is by integrating the simulation with an educational EHR system, such as the one developed by Borycki et al. [14].

Limitations and Directions for Future Research

The main limitations of this pilot study are that we used a small convenience sample and relied on self-reported measures. Although the results are promising, they should be taken with caution as the possibility of type 1 errors cannot be excluded. The intervention was short and only two simulation prototypes were tested, which may explain the lower score for the overall evaluation of the session. For future research, more extensive interventions should be developed and evaluated. A randomized controlled trial with a larger sample and using more objective measures could be implemented, such as video observation of real-life consultations and rating of patient-physician-EMR skills by external observers.

Scale reliability for attitudes related to using the EMR in the consultation was low. This may be due to the small sample

size and the fact that items are substantially different from each other that they may not be unidimensional [15]. To overcome this limitation, we summarized the scores for all items to create a combined attitudes score. However, future research may seek to develop a more reliable instrument for measuring participants' attitudes related to using the EMR in the consultation.

Our goal was to develop a simple, low cost, simulation using tools that do not require expertise in programming or graphic design [6]. However, some participants suggested a more realistic representation of the patients may be useful. Alternative designs that include animation or video segments may better portray patient behaviors—including non-verbal communication—and should be explored.

Finally, the simulation focused exclusively on the use of an EMR system. Increased use of more comprehensive EHRs and patients' access to their information through portals and personal health records (PHRs) introduce new challenges to the patient-clinician interaction during the patient visit. These challenges should be further explored and addressed in future designs of the simulation.

Conclusion

The study suggests that computer-based simulation may be an effective and acceptable tool for teaching Family Medicine residents how to better use the EMR in the consultation, and potentially other clinicians who face similar challenges of using EMRs in the clinical encounter. This potential should be further explored in future research. Usability testing with both clinicians and non-clinicians is a useful approach for identifying a variety of human-computer interaction and scenario design issues.

Acknowledgments

The study was funded by SIM-one (Ontario Simulation Network) Research and Innovation grant. We would like to thank the students and residents who participated in the study. The help of Samer Elamrousy in developing the computer-based simulation is greatly appreciated. Paul Fisher also assisted with the pre-pilot. Professor Joan Cherry was the instructor for the course in which the pre-pilot was held.

References

[1] Toll E. A piece of my mind. The cost of technology. JAMA 2012: 307 (23): 2497-8.

[2] Street RL, Jr., Makoul G, Arora NK, and Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009: 74 (3): 295-301.

[3] Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: a narrative literature review. J Eval Clin Pract 2009: 15 (4): 641-9.

[4] Association of Faculties of Medicine of Canada (AFMC), Canada Health Infoway [Internet]. Environmental scan of e-Health in Canadian undergraduate medical curriculum. 2012 [Cited June 25, 2014]. p. 67. Available from: https://chec-cesc.afmc.ca/sites/all/files/resource_files/e-Health%20Environmental%20Scan_EN_0.pdf

[5] Reis S, Sagi D, Eisenberg O, Kuchnir Y, Azuri J, Shalev V, et al. The impact of residents' training in electronic medical record (EMR) use on their competence: report of a pragmatic trial. Patient Educ Couns 2013: 93(3): 515-21.

[6] Elamrousy S, Domb S, Borycki E, Reis S, Kushniruk A, Ziv A, et al. EMR-sim: a computer-based simulation for enhancing residents' competence in computerized primary care settings. Poster session presented at: Workshop on Interactive Systems in Healthcare (WISH); 2013 November 16; Washington, DC.

[7] Shachak A, Hadas-Dayagi M, Ziv A, and Reis S. Primary care physicians' use of an electronic medical record system: A cognitive task analysis. J Gen Intern Med 2009: 24 (3): 341-8.

[8] Pearce C, Arnold M, Phillips C, Trumble S, Dwan K. The patient and the computer in the primary care consultation. J Am Med Inform Assoc 2011: 18(2): 138-42.

[9] Kushniruk AW, and Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004: 37 (1): 56-76.

[10] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989: 13 (3): 319-40.

[11] Venkatesh V, Morris MG, Gordon BD, and Davis FD. User acceptance of information technology: toward a Unified View. MIS Q 2003: 27 (3) :425-78.

[12] James LR, Demaree RG, and Wolf G. Estimating within-group interrater reliability with and without response bias. J Appl Psychol 1984: 69 (1): 85-98.

[13] Sullivan F, and Wyatt JC. How computers can help to share understanding with patients. BMJ. 2005: 331 (7521): 892-4.

[14] Borycki EM, Kushniruk AW, Joe R, Armstrong B, Otto T, Ho K, et al. The University of Victoria interdisciplinary electronic health record educational portal. Stud Health Technol Inform 2009: 143: 49-54.

[15] Tavakol M, and Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011: 2: 53-5.

Address for correspondence

Dr. Aviv Shachak

Institute of Health Policy, Management and Evaluation, University of Toronto, 155 College St., Toronto, ON M5T 3M6

Phone: 1-416-978-0998, Fax: 1-416-978-7350

Email: aviv.shachak@utoronto.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-511

Mind the Gap: The Discrepancies between Patient Self-Reported Quality of Life and Medical Staff-Estimated Quality of Life

Yumiko Shimamoto[a], Mai Miyabe[a], Shuko Shikata[a], Eiji Aramaki[a]

[a] *Center for the Promotion of Interdisciplinary Education and Research, Kyoto University*

Abstract

Information on patient quality of life (QOL) is essential to many clinical decisions. Therefore, studies that aim to extract QOL information from patient narratives are increasingly drawing attention. Also, several studies have noted that web services for patients, such as patient social networking services, may represent promising resources for QOL research. However, it is still unclear whether patient narrative text contains corresponding amounts of QOL information as self-reported QOL. This study investigates if medical staff can accurately estimate patient QOL from only patient narrative texts. We analyzed (1) QOL of cancer patients estimated by medical staff from patient autobiographical texts and (2) self-reported QOL scores of cancer patients. We compared patients from the following 3 disease groups: (1) gastrointestinal cancer, (2) breast cancer, and (3) lymphoma. The SF-36v2™ Health Survey was used to measure patient QOL in both materials, and the QOLs were compared. We found significant differences between self-reported QOL and estimated QOL in breast cancer patients and lymphoma patients, but not in gastrointestinal cancer patients. In particular, the medical staff tended to underestimate physical QOL scores. Medical staff may underestimate several aspects of QOL scores. On the basis of these results, we may be able to achieve more precise QOL estimation from patient narratives.

Keywords:

Quality of Life; Cancer; Self Report; Narrative Medicine; Social Network Service; Support Systems.

Introduction

Advancements in information and communication technology (ICT) have facilitated the collection of patient narrative data in the form of self-descriptions and self-reports, and the enormous amount of such information (so-called Big Data) is now actively used in medical analyses [1].

"***PatientsLikeMe***"[1] is one of the more successful examples of the use of ICT services in healthcare. In this network, people are able to connect with others with the same disease or condition and share their experiences. In addition to ***PatientsLikeMe***, various other communication services for patients have been increasingly launched over the years [2-5]. "***LifePalette***" is one of the earliest patient communication services established in Japan, and focuses mainly on cancer patients. Another service, "***DIPEx-Japan***", provides patient narratives via videos of patient interviews. Similarly, the "***Healthcare Information Bookshelf Project***" also shares the same goal through the provision of illness narrative books [6]. ICT services are available not only for common diseases, as several rare disease communities have also launched their own social network services (SNS) to communicate with one another. These include the ***GISTERS*** network for gastrointestinal stromal tumor (GIST) patients, the ***Remudy*** registry for dystrophy patients, and the ***Re:me*** network for general rare disease patients. One of the main purposes of such services is to allow patients to share their experiences and support one another. However, the possible secondary applications of such data are also drawing attention.

One promising application of patient narrative data is to understand personal experiences of illness, and to enable training medical staff to improve the quality of care [7-11]. Several studies have reported that patient narratives can provide insight into the feelings of patients. Another secondary use of patient narratives is in quality of life (QOL) research [12, 13]. If accurate QOL data can be obtained from these narratives, the rich amount of text in SNS could provide more patient QOL information than was previously available.

This study is a pilot study to compare narrative data and self-reported data from the viewpoints of QOL. We investigate if medical staffs are able to accurately estimate patient QOL from narrative texts. To investigate this preliminary question, we collected (1) QOL of cancer patients estimated by medical staff based on patient autobiography texts and (2) self-reported QOL scores of cancer patients.

Materials

We utilized 2 types of materials as data sources: Material A (estimated QOL) and Material B (self-reported QOL). In both materials, we used the SF-36v2™ Health Survey (Japanese version) to estimate QOL[2].

Material A (Estimated QOL)

We identified illness narrative books written in Japanese by patients and had been published prior to June 2013. The initial sample comprised 53 books that were available at July 2013 (Authors: 23 men, 30 women; mean age: 36.7±19.7 years). From each of these books, 2 sets of 10 consecutive pages were randomly selected (10*2*53=1,060 pages in total).

These books were classified according to the ICD-10 disease codes for each author (Table 1). QOL estimations were conducted by 5 professional medical staff, comprising of 3 men and 2 women (mean age: 25.6±1.82 years). Each medical staff member read the selected passages, and then estimated the QOL of each patient author based on these passages. The estimations were conducted using the SF-36v2™ Health Survey, which consists of 36 questions. Using information from the books, the medical staff answered the questionnaire by speculating on the author's perspective.

[1] http://www.patientslikeme.com/

[2] This study does not deal with human subjects.

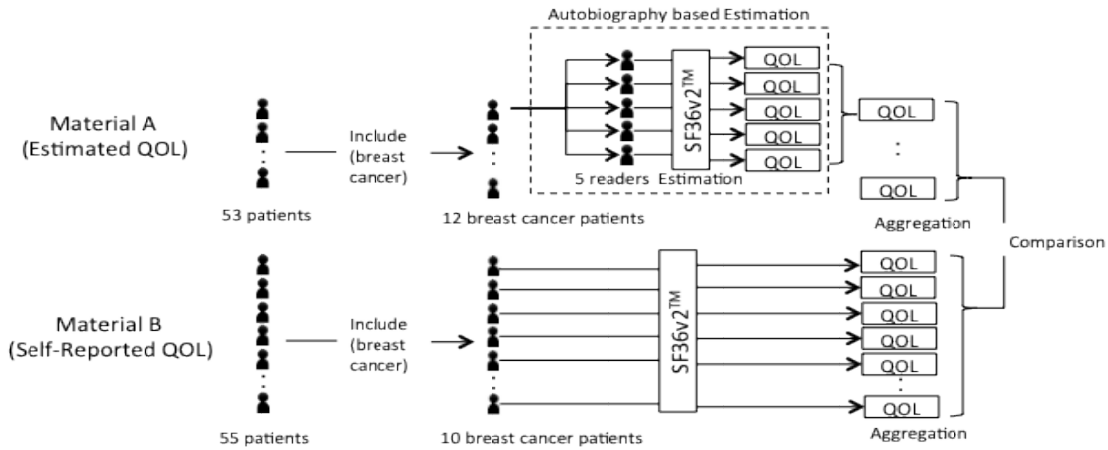


Figure 1– Workflow design of the study in the case of breast cancer patients (ICD-10: C50-58)

Table 1– Material A: Patient Characteristics of the Illness Narrative Book Authors

		n = 53
Age (years)		36.7±19.7
Sex	Male	23 (43.4%)
	Women	30 (56.6%)
ICD-10 Codes	(C00-C14)	3 (5.5%)
	(C15-C26)	11 (20.0%)
	(C30-C39)	6 (10.9%)
	(C40-C41)	1 (1.8%)
	(C43-C44)	0
	(C45-C49)	2 (3.6%)
	(C50-C58)	12 (21.8%)
	(C60-C63)	1 (1.8%)
	(C64-C68)	1 (1.8%)
	(C69-C72)	1 (1.8%)
	(C73-C75)	0
	(C76-C80)	2 (3.6%)
	(C81-C96)	15 (27.3%)

Material B (Self-Reported QOL)

In order to obtain self-reported patient QOL scores, which is the gold standard, we asked 56 patients from ***LifePalette***[3] (an SNS for Japanese cancer patients) to participate in the SF-36v2™ Health Survey. The results were classified according to the ICD-10 codes of each respondent's disease (Table 2). This survey was conducted online through a dedicated website; the evaluation period was from June 16 to June 27, 2014. The effective response rate was 98.2% (13 men, 28 women, and 14 respondents of unknown sex; mean age: 52±9.9 years).

The SF-36v2™ Health Survey (Japanese Ver.)

For the investigation of both materials, we used the SF-36v2™ Health Survey (Japanese ver.)[4] to evaluate the health-related QOL of the cancer patients. The Japanese version of the SF-36v2™ questionnaire produces results in 8 sub-scales and 3 component summary scores (Table 3) [14-16].

Table 2– Material B: Patient Characteristics of the "LifePalette" Users

		n = 55
Age (years)		52±9.9
Sex	Male	13 (24%)
	Women	28 (51%)
	Unknown	14 (25%)
ICD-10 Codes	(C00-C14)	1 (1.8%)
	(C15-C26)	4 (7.1%)
	(C30-C39)	1 (1.8%)
	(C40-C41)	0
	(C43-C44)	0
	(C45-C49)	1 (1.8%)
	(C50-C58)	10 (17.9%)
	(C60-C63)	1 (1.8%)
	(C64-C68)	0
	(C69-C72)	0
	(C73-C75)	1 (1.8%)
	(C76-C80)	2 (3.6%)
	(C81-C96)	4 (7.1%)
	(C97)	0
	Other disease	12(21.6%)
	Unknown	19 (33.9%)

Table 3– Sub-scales of the SF-36v2™ Health Survey

Sub-scales	PF	Physical Functioning
	RP	Role Physical
	BP	Bodily Pain
	GH	General Health
	VT	Vitality
	SF	Social Functioning
	RE	Role Emotional
	MH	Mental Health
Summary Scores	PCS	Physical Component Summary
	MCS	Mental Component Summary
	RCS	Role-social Component Summary

[3] http://lifepalette.jp/

[4] SF-36v2™ Health Survey © 1992, 2000, 2003; QualityMetric Incorporated, Medical Outcomes Trust and Shunichi Fukuhara. All rights reserved. SF-36® is a registered trademark of Medical Outcomes Trust. (SF-36v2 Standard, Japanese)

Methods

The narrative books were obtained in June 2013. The investigation period for Material A was from August 2013 to September 2013. We compared the estimated QOL (Material A) scores and the self-reported QOL (Material B) scores for each disease. From Material A, each of the 5 medical staff members who read the autobiographies produced a QOL value for each patient, resulting in a total of 5 QOL values for that patient. We calculated the average of these 5 QOLs as the estimated QOL of each patient author.

We then classified the patients according to their ICD-10 code disease categories in both Materials A and B (shown in Tables 1 and 2). We calculated the average QOLs in each category, which were regarded as the category QOL.

Comparisons were conducted for categories that had 4 or more patients (n>=4).

Patients were classified into the following 3 categories according to their ICD-10 codes: (1) C15-26 (Malignant neoplasms, digestive organs; or "gastrointestinal cancer"), (2) C50-58 (Malignant neoplasms, breast and female genital organs; or "breast cancer"), and (3) C81-96 (Malignant neoplasms, stated or presumed to be primary, of lymphoid, hematopoietic and related tissue; or "lymphoma"). Figure 1 illustrates the workflow design of this study using the example of breast cancer patients.

Results

The results are shown in Figure 2 for ICD-10 codes C15-26 (gastrointestinal cancer), Figure 3 for C50-58 (breast cancer), and Figure 4 for C81-96 (lymphoma).

In gastrointestinal cancer patients (Figure 2), the results showed no significant differences between Material A (n = 11) and Material B (n = 4) in all sub-scales.

In breast cancer patients (Figure 3), there were significant differences observed between Material A (n = 12) and Material B (n = 10) in 3 of the sub-scales (PF, RP, and GH) and one of the 3 summary scales (PCS).

Lymphoma patients (Figure 4) showed smilar results to the breast cancer patients (Figure 2): there were significant differences observed between Material A (n = 15) and Material B (n =4) in 4 of the sub-scales (PF, RP, VP, and SF) and one of the 3 summary scales (PCS).

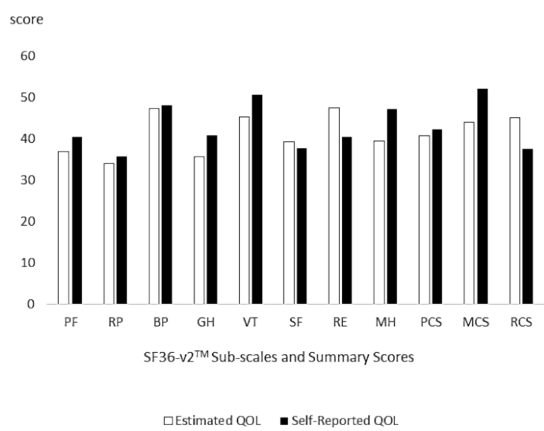


Figure 2– C15-26 (Differences between estimated QOL and self-reported QOL in gastrointestinal cancer patients)

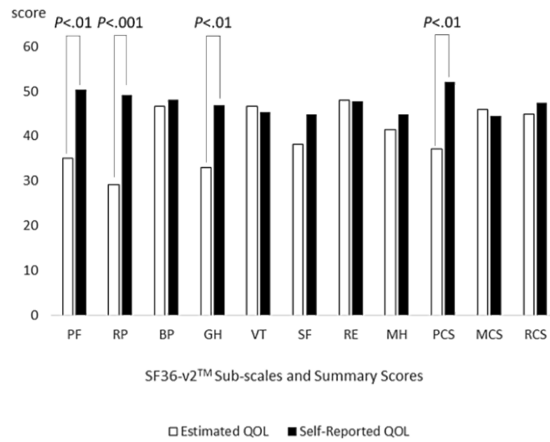


Figure 3– C50-58 (Differences between estimated QOL and self-reported QOL in breast cancer patients)

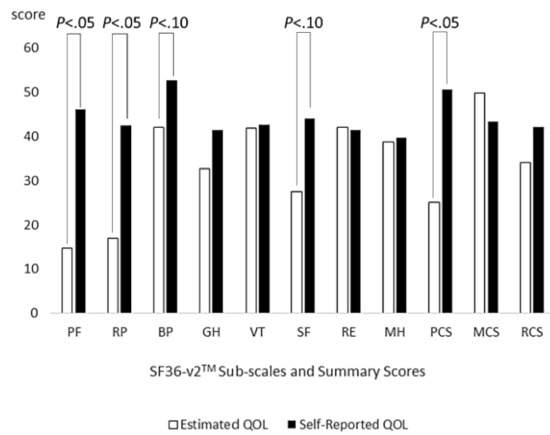


Figure 4– C81-96 (Differences between estimated QOL and self-reported QOL in lymphoma patients)

Discussion

Our findings showed that among the 3 summary scores of the SF-36v2^{TM} Health Survey, there were statistically significant differences in the Physical Component Summary (PCS) score between medical staff-estimated QOL and self-reported QOL in 2 of the 3 types of cancers analyzed. Although various studies have addressed patients suffering from these diseases, our study, surprisingly, indicated that the medical staff tended to underestimate patients' physical QOL. This difference in PCS scores between the medical staff and patient authors suggests that patients may tend to speak out explicitly regarding their pain. If so, patients may suffer from less pain than expected by the medical staff. In the near future, we would have more precise techniques to estimate the level of patient pain.

The other QOL summary scores (mental component and social component) did not show any significant differences between medical staff estimations and self-reporting. This suggests that the patient narratives contain much more information on mental and social activities, thereby enabling the medical staff to have a better interpretation of these aspects of QOL. As most patient narratives focus on their daily lives, it is reasonable that the medical staff were able to accurately estimate the mental and social aspects of patient QOL. In most cases, the QOL gap was due to underestimations by medical staff. In other words, medical staff rarely overestimated the patient

QOL scores. One of the possible reasons for this bias is that the medical staff may have been too intent on empathizing with the patients and sharing their feelings, which resulted in the gap.

Among the 3 cancer groups, the most precise QOL estimation was achieved in gastrointestinal cancer (Figure 2). In Japan, the mortality risk of gastrointestinal cancer is the second highest after lung cancer. This study showed that even in such a serious disease, there was little difference between the medical staff-estimated QOL and patient self-reported QOL; this suggests a degree of success in information sharing for this disease. In contrast, there were gaps between the 2 methods of QOL assessment in the other 2 cancer groups.

The results for breast cancer were not unexpected, as breast cancer has a relatively low mortality risk. In contrast, the mortality risk of lymphoma is higher than many other types of cancers. The reasons why the medical staff tended to underestimate the QOL for this disease require further study.

A surprising finding was that the self-reported QOL scores tended to be relatively high. For example, the Bodily Pain (BP) sub-scale in lymphoma patients was higher than the Japanese average (in this QOL scale, the average Japanese person has a score of 50 in each of the QOL scale components). In addition to the BP value, most of the other QOL sub-scales were over 40. This result suggests that cancer patients may be happier than we had expected.

Limitations

A limitation of this study is that the 2 aspects of the analysis were obtained from different media, which may have biased the results: Material A was acquired from autobiographical books and Material B from patients through SNS. However, we analyzed the same diseases in order to reduce the possible bias.

Conclusions

In this study, we compared medical staff-estimated QOL and patient self-reported QOL scores in 3 categories of cancer patients. Our findings demonstrate that the medical staff tended to underestimate the physical QOL of patients based on autobiographical text. In contrast, there were no significant differences between the 2 types of QOL assessments in the mental and social aspects. Care should therefore be taken when estimating QOL from illness narratives, particularly for the physical component of QOL. These results indicate a need to create support systems that can improve patient conditions by using the gaps in physical QOL assessment. In the future, if a technique for automated QOL estimation is developed, the result of this study will contribute to effective communication between patients and medical professionals.

Acknowledgments

This study was supported in part by JST PRESTO. The authors wish to thank Mediaid Corporation for providing blog articles on the SNS “LifePalette.”

References

[1] Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013: 346(f167): 1-5.

[2] Wada E. Narrative Clues of Illness in Published Autobiographical Documents. [in Japanese]. Journal of St. Luke's Society of Nursing Research 2003: 7(1): 1-8.

[3] Tsukuda S and Okawa I. Positive Changes After Cancer Experience Observed in Illness Records [in Japanese]. Japanese Journal of Clinical Psychology 2013: 13(6): 839-48.

[4] Shimamoto Y, Kawashima K, Usuda Y, Aramaki E. The Clinical Utility of Humor in Illness: The Relationship Between Humor and Quality of Life as Shown in Cancer Patient Narratives [in Japanese]. Journal of the Japan Medical Library 2014: 61(3): 499-508.

[5] Wada E. What will the "Tobyo-ki Collection" bring to patients and medical professionals? Experience through professional collaboration of "Healthcare Information Bookshelf Project": Experience through professional collaboration of "Healthcare Information Bookshelf Project" [in Japanese]. Journal of Information Processing and Management 2006: 49(9):499-508.

[6] Abe Y. Qualitative Study Examining What Medical Staff Obtain by Reading Illness Narratives (Tobyo-ki) [in Japanese]. Journal of the Japan Medical Library 2012: 59(3):176-9.

[7] Wada E. A Reading Circle with Nurses in Regard to Autobiographical Illness Narratives. [in Japanese]. Pharmaceutical library bulletin 2011: 56(3): 235-9.

[8] Kodaira T, Ito T, Ohtaka Y. A Text Mining Analysis of an Autobiographical Illness Narrative Book : "What is the Meaning of Mental Disease?: The Experience of a Patient with Mental Illness" Written by Natsuko Furukawa [in Japanese]. Journal of Japan Academy of Psychiatric and Mental Health Nursing 2011: 19(2): 10-21.

[9] Donzelli G, Paddeu EM, D'Alessandro F, Costa AN. The role of narrative medicine in pregnancy after liver transplantation. The journal of maternal-fetal & neonatal medicine, 2014; pp. 1-4.

[10] Johna S, Woodward B, Patel S. What can we learn from narratives in medical education? The Permanente Journal 2014: 18(2): 92-4.

[11] Egerod I, Christensen D, Schwartz-Nielsen KH, Ågård AS. Constructing the illness narrative: A grounded theory exploring patients’ and relatives’ use of intensive care diaries. Critical Care Medicine 2011: 39(8): 1922-8.

[12] Sakellariou D, Boniface G, Brown P. Using Joint Interviews in a Narrative- Based Study on Illness Experiences. Qualitative Health Research 2013: 23(11): 1563-70.

[13] Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. Journal of Clinical Epidemiology 1998: 51(11): 1037-44.

[14] Fukuhara S, Ware JE, Kosinski M, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey, Journal of Clinical Epidemiology 1998: 51(11): 1045-53.

[15] Fukuhara S and Suzukamo Y. Manual of SF-36v2 Japanese version, Institute for Health Outcomes & Process Evaluation research. Kyoto, 2004.

[16] Center for Cancer Control and Information Services, National Cancer Center, Japan. [http://ganjoho.jp/professional/statistics/statistics.html#mortality] (December 2014)

Address for correspondence

Eiji ARAMAKI <eiji.aramaki@gmail.com>
Center for the Promotion of Interdisciplinary Education and Research, Kyoto University. Kyoto 600-8815, Japan.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-515

Identifying Effective Approaches for Dissemination of Clinical Evidence – Correlation Analyses on Promotional Activities and Usage of a Guideline-Driven Interactive Case Simulation Tool in a Statewide HIV-HCV-STD Clinical Education Program

Dongwen Wang[a], Xuan Hung Le[a], Amneris E. Luque[a]

[a] *University of Rochester Medical Center, Rochester, NY, USA*

Abstract

Dissemination of the latest clinical evidence to community-based healthcare providers is a critical step to translate biomedical knowledge into clinical practice. We performed a study to analyze the correlations between the promotional activities and the usage of a guideline-driven interactive case simulation tool (ICST) for insomnia screening and treatment in a statewide HIV-HCV-STD clinical education program. For this purpose, we tracked users' interactions with the ICST and the sending of promotional email newsletters during a study period of 44 weeks. Results showed that promotional activities were strongly correlated with the number of audience as well as the intensity of use of the target resource. The strength of correlation varied in specific use contexts. Strong correlations were found between the sending of email newsletters and the intensity of resource use by promotion recipients, by new users, and through the most convenient access channel associated with the promotion. Selection of approaches for resource dissemination should consider the potentials and limitations of use contexts to make them more effective.

Keywords:

Information Dissemination; Clinical Practice Guideline; Evidence-Based Medicine; Interactive Case Simulation Tool; Insomnia; HIV; HCV; STD.

Introduction

Effective and timely dissemination of the latest clinical evidence to community-based healthcare providers is a critical step to translate biomedical knowledge into clinical practice. Dissemination and implementation research has become a national priority of the United States to harness the investment in biomedical research with the improvement in clinical care and public health [1]. The informatics research community has accumulated a critical mass of knowledge on effective approaches to disseminating clinical evidence, including use of online platforms to host clinical and educational resources and leverage of information and telecommunication technologies to promote them [2-3]. Nevertheless, few prior studies have reported on how promotional activities can improve the actual usage of online resources, and particularly, in what specific contexts the resource usage tends to respond to promotions.

In this paper, we report a pilot study to analyze the correlations between the usage of a guideline-driven interactive case simulation tool (ICST) and resource dissemination activities [4-6]. The results from this study will: (1) identify how dissemination activities can increase the usage of online clinical and educational resources; (2) characterize the profiles of use contexts that respond to promotions; and (3) direct the future development of effective approaches to disseminating clinical evidence.

Materials and Methods

Target Resources

The target resources for dissemination in this study were from the New York State (NYS) HIV-HCV-STD Clinical Education Initiative (CEI) online training program [7]. The CEI program is sponsored by NYS Department of Health (DOH) AIDS Institute, targeting primary care clinicians such as physicians, nurse practitioners, pharmacists, dentists, physician assistants, dental hygienists, case managers, social workers, counselors, and other team members providing care for HIV-HCV-STD patients. CEI aims to increase access to quality HIV-HCV-STD healthcare, to expand the base of clinicians who can effectively manage HIV-HCV-STD patients, to disseminate the latest clinical guidelines, and to foster partnerships between community-based care providers and HIV-HCV-STD specialists. Since 2008, CEI has initiated the online training program and developed various digital resources, including 235 multimedia learning modules, 93 online CME/CNE courses, and 12 guideline-driven ICSTs [8-9]. These resources can be accessed from multiple platforms, including a main website, a mobile website, Android and iOS mobile apps, online social networks and social media, RSS feeds, and newsletters sent through CEI email listserv. Over the five-year period since its launch, the CEI website has recorded 121,150 visits and 608,487 pageviews by audiences from 171 countries around the world [6,10]. It is now consistently ranked by Google and other search engines as a top site for HIV-HCV-STD clinical education.

ICST

ICSTs are online tools for clinical decision assistance and case simulation on individual patients [4, 11]. When used by healthcare providers, they serve as vehicles to dessiminate the latest biomedical knowledge. On the back-end, ICSTs are typically based on knowledge-bases adopted from clinical practice guidelines. For example, the *Insomnia Screening and Treatment* ICST selected as the resource for analyses in this study was based on a quick reference guide for HIV primary care clinicians, a simplified version of the full guideline developed by the NYS DOH AIDS Institute [4, 12]. On the front-end, ICSTs support user interactions such as review of patient management processes, examination of different

options for clinical decisions, and entry of case-specific data for individualized recommendations. In addition to user-defined patient cases, an ICST can include a list of pre-defined sample cases presenting the typical scenarios for clinical decision making as well as a text recommendation section with highlighted points adopted from the source guideline. These three sections, i.e., *recommendation*, *sample case*, and *user-defined case*, provide the complete user functions for an ICST [4]. The ICSTs can be accessed from the CEI website as well as mobile apps. Screenshots of the *Insomnia Screening and Treatment* ICST are shown in Figure 1.

CEI Resource Promotion

To ensure that its resources can reach out to the target clinicians and communities, CEI has been consistently engaging in program promotions through multiple channels. Among these, CEI newsletters sent to email listserv have been used most freqently. The CEI newsletters are regular updates of the latest CEI clinical and educational resources, such as the newly developed CME/CNE courses, multimedia learning modules, mobile tools, training events and news, etc. These newsletters contain hyperlinks pointing to the CEI website and other online platforms that host the various clinical and educational resources. For many years, we have been sending these newsletters to the CEI audience through email listserv every two to three weeks. The number of subscribers of this listserv is in the range of 2,200-2,600. When a listserv subscriber receives a CEI newsletter and finds interest in a specific item, he/she can click the hyperlink to check the details of that resource on the CEI website. He/she can also forward the email to colleagues or friends to further disseminate these resources through his/her professional and social networks [6, 13].

Study Design and Data Collection

To analyze the correlations between the ICST usage and promotional activities, we first collected the usage data from April 3, 2012 (when the *Insomnia Screening and Treatment* ICST was released) to February 4, 2013 (a period of 44 weeks) through tracking of user interactions with the system [4]. We focused on two usage measures: (1) episodes of use (reflecting the number of audience); and (2) frequency of visits to specific part of the ICST (reflecting the intensity of use) [4]. For this purpose, we built customized tracking functions to monitor the use of specific functions on different system user interfaces. System and user events, such as click of a button, check of a decision branch, and selection of a specific patient condition, as well as the timestamps of these events and user sessions, were stored in a tracking database for analyses. To control potential biases, usage by internal staff and for testing of app installation was excluded from the analyses [4].

To characterize the use context, we profiled user interactions by multiple dimensions:

1. Sections of ICST system function, i.e., *user-defined case*, *sample case*, *recommendation*, and *cross-box* (usage spanning over two or three of the previously listed sections);
2. *New* vs. *returned users*;
3. Access from *large-screen equipment* (desktops, tablets, etc.) vs. *small-screen hand-held devices* (smartphones, iPods, etc.);
4. Access through *web browsers* vs. *native apps*;
5. Audience from the *United States* vs. *non-US countries*;
6. Audience from the *NYS* vs. *out of the state*; and
7. Audience from *healthcare organizations*, *government agencies*, and *unknown settings*.

Additional technical details on tracking of user interactions, use contexts, and usage measures can be found in our previous publications [4, 13-14].

To collect the data of CEI resource dissemination through newsletters, we reviewed all the archived emails, selected those newsletters directly related to ICSTs, and recorded the date when they were sent.

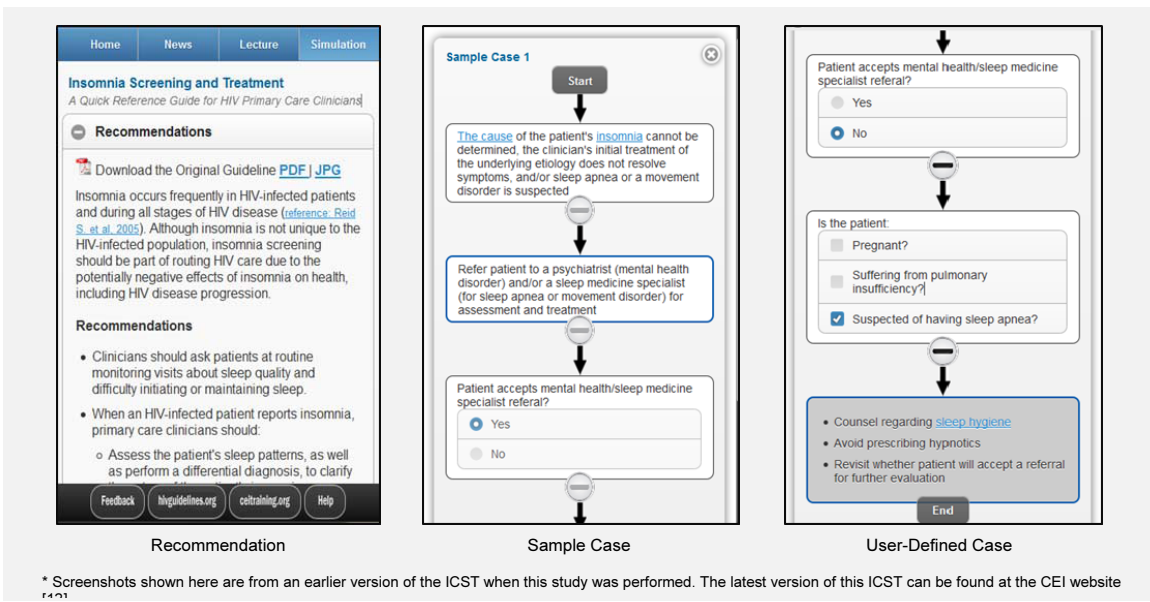


Figure 1 – Screenshots of the Insomnia Screening and Treatment ICST

Data Analyses

For data analyses, we segmented the total usage as well as the profiles of use context by the dimensions listed above as bi-weekly time-series. We plotted dissemination activities, i.e., sending CEI newsletters through emails, as a binary variable in a bi-weekly time-series with the same formulation. We then performed correlation analyses between each combination of usage profile/measure and the dissemination activities. We calculated Pearson correlation coefficient (r) to determine the potential correlations. We considered $|r| >= 0.5$ as a strong correlation, $0.5 > |r| >= 0.3$ as a moderate correlation, and $|r| < 0.3$ as a weak or no correlation [15]. For statistical analyeses, we used the SPSS software package [16].

Results

A total of 298 episodes and 1,415 rounds of interactions (visits) with the *Insomnia Screening and Treatment* ICST were recorded during the study period. Meanwhile, five CEI email newsletters were sent during this period to promote the ICSTs. The detailed distributions of the total ICST usage, its use in specific contexts, and the sending of email newsletters presented as bi-weekly time-series are shown in Figure 2.

Table 1 – Results of Correlation Analyses

	Episodes		Freq. of Visits	
	r	p-value	*r*	p-value
Recommendation	0.470	0.024	0.551	0.006
Sample Case	-0.091	0.679	0.224	0.303
User-Defined Case	-0.066	0.764	0.270	0.213
Cross-Box	0.278	0.198	0.437	0.037
New User	0.492	0.017	0.538	0.008
Returned User	0.446	0.033	0.122	0.579
Large Screen	0.422	0.045	0.486	0.019
Small Screen	0.282	0.192	0.271	0.211
Web Browser	0.460	0.027	0.585	0.003
Native App	0.113	0.607	0.173	0.429
Outside of USA	0.284	0.189	0.219	0.315
Inside of USA	0.474	0.022	0.502	0.015
Outside of NYS	0.486	0.019	0.484	0.019
Inside of NYS	0.452	0.030	0.342	0.110
Healthcare Organization	0.462	0.027	0.607	0.002
Government Agency	0.420	0.046	0.548	0.007
Unknown Settings	0.463	0.026	0.325	0.131
Total Usage	0.500	0.015	0.484	0.019

Note: *r*: Pearson correlation coefficient; dark grey cells: strong correlation ($|r| >= 0.5$); light grey cells: moderate correlction ($0.5 > |r| >= 0.3$); white cells: weak or no correlation ($|r| < 0.3$).

Pearson correlation coefficients showed strong correlations between promotional activities and the following usage profiles:

1. Usage measured by episodes on *total usage*; and
2. Usage measured by frequency of visits on *recommendation*, by *new users*, through *web browsers*, by audience from *the United States*, and by audience from *healthcare organizations* and *government agencies*.

Meanwhile, Pearson correlation coefficients showed that there were moderate correlations between promotional activities and the following usage profiles:

1. Usage measured by episodes on *recommendation*, by both *new* and *returned users*, from *large-screen equipment*, through *web browsers*, by audience from *the United States*, by audience from both *inside* and *outside of NYS*, and by audience from *all settings*; and
2. Usage measured by freqency of visits from *large-screen equipment*, by audiences from both *inside* and *outside of NYS*, and on *total usage*.

Weak or no correlations were found between promotional activities and the following usage profiles:

1. Usage measured by episodes on *sample case* and *user-defined case*, from *small-screen hand-held devices*, through *native apps*, and by audience from *non-US countries*; and
2. Usage measured by freqency of visits on *sample case* and *user-defined case*, from *returned users*, from *small-screen hand-held devices*, through *native apps*, and by audience from *non-US countries*.

The detailed results of the correlation analyses, including *r* and p-value for each usage profile, are summarized in Table 1.

Discussions

The results have confirmed that there were moderate (measured by intensity of use) to strong (measured by number of audience) correlations between the *total usage* of the *Insomnia Screening and Treatment* ICST and the sending of CEI email newsletters. The strength of correlation in specific use contexts was not evenly distributed. Instead, certain usage profiles demonstrated strong or moderate correlations, while others only presented weak or no correlations. This is the first time a study has characterized the profiles of use context that do and do not respond to promotions. These identified usage profiles will direct future informatics research to develop effective and targeted approaches to disseminate clinical evidence, focusing on particular users, platforms, and settings.

Regarding the specific use contexts, here are the potential explanations on strength of the correlations:

1. The *recommendation* section of the ICST system function was the target of the promotional email (clicking the ICST hyperlink in the email would bring a user to this section). Therefore, its use demonstrated a strong correlation with the promotional emails. In contrast, the *sample case* and *user-defined case* sections presented only weak or no correlations.
2. The *new users* had not yet bookmarked the ICST webpage nor downloaded the mobile apps. Therefore, they were more likely to respond to email promotions.
3. Access through *web browsers* was the default channel when clicking the ICST hyperlink in the promotional email. Therefore, this use context recorded more responses. In contrast, access through *native apps* and *small-screen hand-held devices* (hosting the native apps) only showed weak or no correlations.
4. Most of the CEI newsletter subscribers were from *the United States*. Therefore, these audiences were more likely to respond to email promotions.
5. Most of the CEI program audiences were clinicians and public health professionals. Therefore, the audiences from *healthcare organizations* and *government agencies* demonstrated strong correlations.

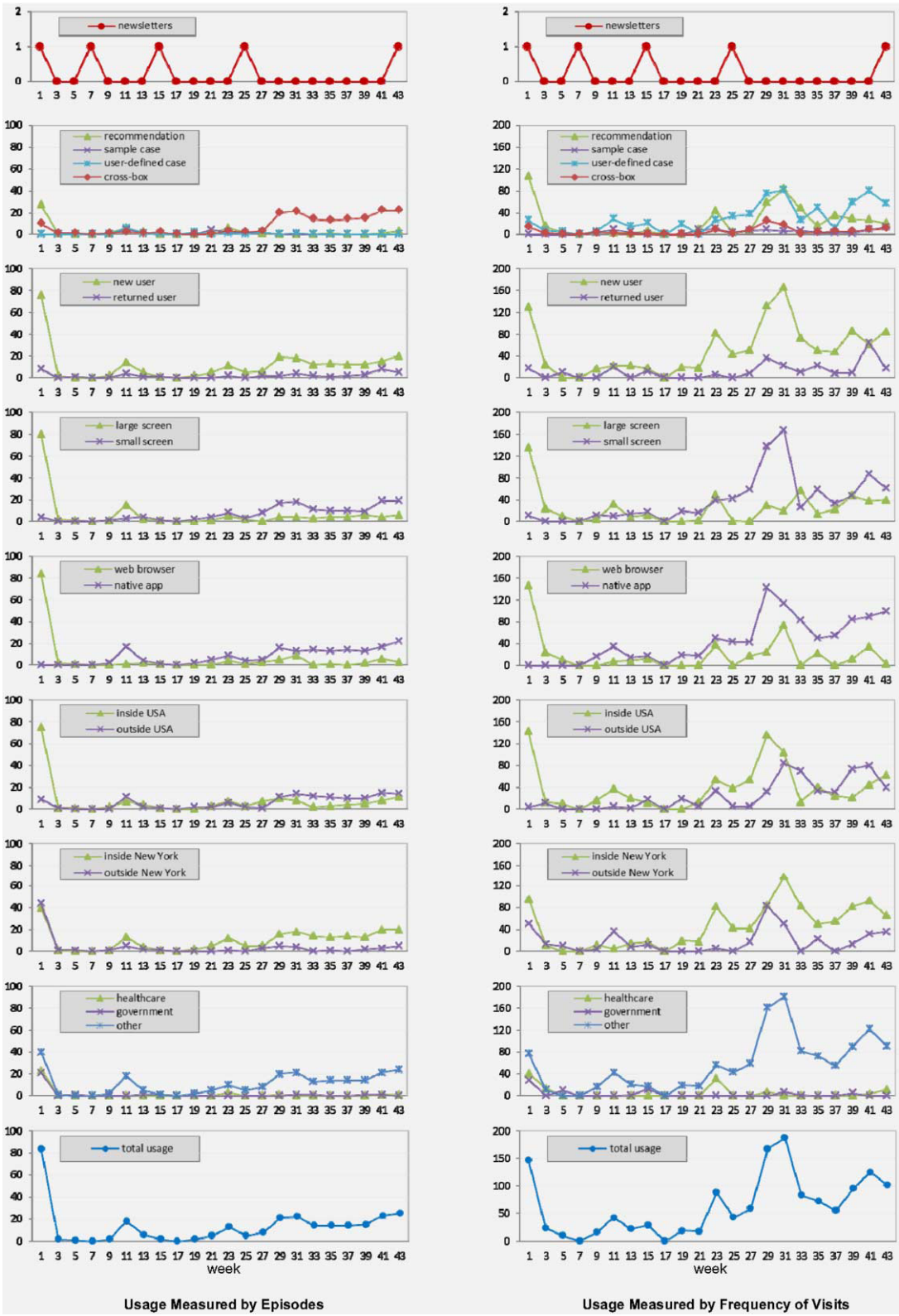


Figure 2 – Sending of CEI Email Newsletters, Total ICST Usage, and ICST Use in Specific Contexts during the Study Period

To plan for effective resource dissemination, the use contexts with strong correlations should be the focus. Regarding the use contexts demonstrating only weak or no correlations, resource usage is likely from a core group of loyal audiences (considering that we still recorded high level of usage in these use contexts) who do not need promotions.

There were two limitations in this study. First, we selected a single ICST as the target resource and only the email newsletter as the dissemination activity for analyses. There were many other resources developed by the CEI online training program, and these resources were promoted through multiple channels. Generalizability of the findings from this study, therefore, needs to be verified in future research. Second, this study focused only on correlations, which could be the first step to determine causality. For the use contexts with strong correlations, it may worth further investigating on the detailed pathways in which the dissemination activities can lead to the increasing use of clinical and educational resources.

Conclusion

Dissemination of clinical evidence through promotional activities is correlated with number of audience and intensity of use of online education resources. Strength of correlation depends on specific use contexts. Strong correlations were found between sending of email newsletters and intensity of resource use by promotion recipients, new users, and through the most convenient access channel associated with the promotion. Selection of approaches for resource dissemination should consider their potentials and limitations in specific use contexts to make them more effective.

Acknowledgments

This work is supported by NYS DOH AIDS Institute through contracts C023557, C024882, and C029086, and by the Agency for Healthcare Research and Quality (AHRQ) through grant R24 HS022057. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors. We would like to thank: 1) CEI staff Monica Barbosu, Matthew Bernhardt, and Terry Doll for their contributions to the study; and 2) NYS and AHRQ program officers Cheryl Smith, Howard Lavigne, Beatrice Aladin, Lyn Stevens, Bruce Agins, and Marian James for their support.

References

[1] Zerhouni EA. Translational and clinical science – time for a new vision. N Engl J Med 2005: 353(15): 1621-3.

[2] Schnall R, Cimino JJ, and Bakken S. Development of a prototype continuity of care record with context-specific links to meet the information needs of case managers for persons living with HIV. Int J Med Inform 2012: 81(8): 549-55.

[3] Del Fiol G, Curtis C, Cimino JJ, Iskander A, Kalluri AS, Jing X, Hulse NC, Long J, Overby CL, Schardt C, and Douglas DM. Disseminating context-specific access to online knowledge resources within electronic health record systems. Stud Health Technol Inform 2013: 192: 672-6.

[4] Le XH, Luque A, and Wang D. Assessing the usage of a guideline-driven interactive case simulation tool for insomnia screening and treatment in an HIV clinical education program. Stud Health Technol Inform 2013: 192: 323-7.

[5] Wang D, Luque A, and Le XH. Engaging HIV healthcare providers in online learning through interactive case simulation tools. Proc 7th IAS Conference on HIV Pathogenesis, Treatment and Prevention (IAS 2013), 2013; MOPE196.

[6] Le XH, Luque A, and Wang D. Leveraging information technologies and multiple online platforms to disseminate HIV/AIDS clinical evidences to community healthcare providers. Proc 7th Annual Conference on the Science of Dissemination and Implementation: Transforming Health Systems to Optimize Individual and Population Health, 2014.

[7] New York State HIV-HCV-STD Clinical Education Initiative. Available at: http://ceitraining.org. Accessed on Dec. 5, 2014.

[8] Wang D, Le XH, and Luque A. Development of digital repositories of multimedia learning modules and interactive case simulation tools for a statewide clinical education program. Proc 6th International Workshop on Knowledge Representation for Health-Care (KR4HC-2014) 2014: 145-51.

[9] Le XH, Barbosu M, Doll T, Bernhardt M, DellaPorta T, Luque A, and Wang D. Enhancing dissemination of HIV-HCV-STD clinical evidence through cross linkages of multimedia learning modules, interactive case simulation tools, and clinical practice guidelines. Proc Workshop on Interactive Systems in Healthcare (WISH-2014), 2014.

[10] Wang D, Luque A, Le XH, Doll T, Barbosu M, DellaPorta T, Lavigne H, Smith C, and Agins B. Development of an online clinical education program for New York State HIV healthcare providers. Proc XIX International AIDS Conference (AIDS 2012) 2012: 5376.

[11] Le XH, Luque A, and Wang D. Development of guideline-driven mobile applications for clinical education and decision support with customization to individual patient cases. AMIA Annu Symp Proc 2012: 1828.

[12] Insomnia Screening and Treatment Case Simulation. Available at: http://m.ceitraining.org/guidelines/insomnia. Accessed on Dec. 5, 2014.

[13] Le XH, Luque A, and Wang D. Correlations between promotional activities and usage of a guideline-driven interactive case simulation tool. AMIA Annu Symp Proc 2013: 839.

[14] Le XH, Luque A, and Wang D. Analyses of usage patterns of an interactive case simulation tool and its access contexts – a case study of an insomnia screening and treatment module for HIV clinical education. Proc 6th Annual Mid-Atlantic Healthcare Informatics Symposium, 2013.

[15] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum Associates, 1988.

[16] SPSS software. Available at: http://www.ibm.com/software/analytics/spss. Accessed on Dec. 5, 2014.

Address for correspondence

Dongwen Wang, PhD; University of Rochester Medical Center, 601 Elmwood Ave., Box 630, Rochester, NY 14642, USA.
Email: dongwen_wang@urmc.rochester.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-520

Building Comprehensive and Sustainable Health Informatics Institutions in Developing Countries: Moi University Experience

Martin C. Were[a,b,c], Abraham Siika[b,c], Paul O. Ayuo[b,c], Lukoye Atwoli[b], Fabian Esamai[b]

[a] *Regenstrief Institute, Inc. & Indiana University School of Medicine, Indianapolis, IN, USA*
[b] *Moi University School of Medicine, Eldoret, Kenya*
[c] *Academic Model Providing Access to HealthCare (AMPATH), Eldoret, Kenya*

Abstract

Current approaches for capacity building in Health Informatics (HI) in developing countries mostly focus on training, and often rely on support from foreign entities. In this paper, we describe a comprehensive and multi-dimensional capacity-building framework by Lansang & Dennis, and its application for HI capacity building as implemented in a higher-education institution in Kenya. This framework incorporates training, learning-by-doing, partnerships, and centers of excellence. At Moi University (Kenya), the training dimensions include an accredited Masters in HI Program, PhD in HI, and HI short courses. Learning-by-doing occurs through work within MOH facilities at the AMPATH care and treatment program serving 3 million people. Moi University has formed strategic HI partnerships with Regenstrief Institute, Inc. (USA), University of Bergen (Norway), and Makerere University (Uganda), among others. The University has also created an Institute of Biomedical Informatics to serve as an HI Center of Excellence in the region. This Institute has divisions in Training, Research, Service and Administration. The HI capacity-building approach by Moi provides a model for adoption by other institutions in resource-limited settings.

Keywords:

Health Informatics; Capacity Building; Developing Countries; Training.

Introduction

Health Informatics (HI) is rapidly being embraced in the developing world[1]. It is now commonplace to see comprehensive deployment and implementation of Electronic Health Record Systems (EHRs). As an example, countries like Rwanda and Kenya have rolled out EHRs at scale within Ministry of Health (MOH) facilities. Other health information systems, ranging from Health Information Management Systems, Laboratory Information Systems, Pharmacy Information Systems, and mobile Health Systems also have increased uptake. Beyond systems, there are multiple initiatives to develop and operationalize national Health Information System (HIS) strategies. At the same time, increasing numbers of evaluations are underway to increase the evidence-base of impact of HIS.

With increasing relevance of HI in developing countries comes the requisite need for comprehensive mechanisms to support these systems. At the very least, countries need appropriate institutional and human capacity in HI. Needed workforce in HI include: (a) Local level: Health IT professionals, eHealth specialized programmers, data managers, implementation managers, support specialists and reporting personnel; (b) Institutional level: Chief medical information officers; and health information management specialists, (c) Administrative: national and regional eHealth coordinators and eHealth monitoring and evaluation specialists, and (d) Other: Health information privacy and security specialists, HI researchers, among others. End users, managers and policy-makers also need appropriate training on relevant aspects of HI.

Over the years, various approaches have been undertaken to improve capacity in HI for developing countries. There are several examples of training programs. The American Medical Informatics Association has had the 10x10 program whose aim is to train 10,000 health care professionals in applied health and medical informatics within 10 years, with some coming from developing countries[2]. The Informatics Training for Global Health Program of the Fogarty International Center is another initiative that aims to improve informatics research training in developing countries[3]. Health Informatics Building Blocks (HIBBS) by the Global Health Informatics Partnership arm of AMIA also provides applied informatics training targeted to developing countries.

Beyond formalized training programs, other approaches to improving capacity have involved direct presence in developing countries of Western institutions, or significant North-South collaboration. Organizations like the International Training and Education Center for Health (ITECH) conducts extensive work and training in HI in multiple developing countries. Regenstrief Institute, Inc., Partners In Health, and numerous universities in the West now have formal relationships with developing country partners and ministries in HI. An exemplar case of institutional partnerships for HI capacity building involves collaboration between University of Washington, Seattle and Universidad Peruana Cayetano Heredia, Peru[4, 5]. The project, called AMUATA, has lasted for over a decade and led to significantly increased capacity through short-term courses, Masters and PhD-level HI training.

Unfortunately, few HI capacity building initiatives can mimic the success of programs like AMUATA. The problem is that most initiatives target narrow dimensions of a broader capacity building framework. It is quite common for individuals to be trained in HI within the Western setting, only to come back home and have no institutional base from which to operate. In fact, these individuals often receive little requisite recognition for their newly acquired skills. Numerous capacity building initiatives also lack buy-in from relevant governmental authorities, with efforts not necessarily aligned with larger national HI capacity-building goals. Some HI training programs also lack solid implementation partnerships to enable applied training in HI within real-world settings.

In this paper, we report on an HI capacity building initiative grounded by a comprehensive and multi-dimensional capacity building framework. We use this case-study approach to high-

light the necessity of a multi-faceted approach to HI capacity building, with the primary goal of realizing self-sustainable locally run HI programs in developing countries.

Materials and Methods

Capacity Building Framework

Lansang & Dennis describe a broad-based capacity building framework that takes into consideration both short- and long-term strategies for capacity building with an eye towards sustainability (**Table 1**)[6]. This framework incorporates various complementary approaches for capacity building, including: (1) graduate or post graduate training; (2) learning by doing; (3) institutional partnerships between developing and developed countries; and (4) centers of excellence. Building HI capacity at the institutional level requires an approach that touches on every aspect of the four dimensions in this framework. It should be noted that training, the most common modality currently used globally for HI capacity building, leads to a lower likelihood for sustainability if implemented in isolation.

Approach

The HI program described in this paper used the comprehensive framework by Lansang & Dennis to build a broad-based and sustainable HI Informatics Program within an institution in a developing country setting. The guiding principles for the HI program are outlined below:

- *Multi-Dimensionality:* The HI program should support various dimensions of HI, including: HI research, HI system development and implementation, and policy guidance.
- *Locally Led:* The program should aim for self-suffiency of the local institution, with little reliance from outside.
- *Strategic Partnerships:* The program should form public-private, North-South, and South-South strategic partnerships as needed to achieve its goals.
- *Government support and Integration:* The program should ideally operate within publicly-funded institutions, and work closely and in alignment with government-supported health and education sectors.
- *Accreditation:* Training programs should be formally recognized within the country, and follow well-accepted core competencies in HI training and education[7-9].
- *Affiliation with an Implementation Setting:* There needs to be an emphasis on applied and practice-based HI. As such, programs should collaborate closely with active care systems, and with larger national HIS ecosystem.
- *Economic sustainability:* Multi-modal approaches for finanacial sustainability, incorporating various business models, should be implemented to avoid excessive reliance on seasonal and time-limited grants or funding approaches.

Setting

The comprehensive HI program, guided by the Lansang & Dennis framework, has been implemented within Moi University in Kenya. Moi University is a public and government supported university located in Eldoret, Kenya[10]. The University was established in 1984 as the second university in Kenya through an Act of Parliament. It currently has 13 schools, 4 directorates and operates 6 campuses, namely: Main Campus, Eldoret West Campus, Odera Akang'o Campus, Mombasa Campus, Nairobi Campus, and Eldoret Town campus. Within the Town Campus is the College of Health Sciences that houses four schools, namely: School of Medicine, School of Public Health, School of Dentistry and School of Nursing. The College of Health Sciences has a student population of about 2,000 with a staff complement of over 500 of which 170 are academic staff. The College of Health Sciences hosts the HI programs.

Results

We describe the activities of the Moi University HI program along the four dimensions of capacity building framework outlined in the Lansang & Dennis framework, namely: (1) Graduate or post graduate training; (2) Learning by doing; (3) Institutional partnerships between developing and developed countries; and (4) Centers of excellence.

Graduate and Post Graduate Training

With support from the Norwegian Agency for Development Cooperation, and as part of the NORHED program, Moi University is leading the Health Informatics Training and Research in East Africa for Improve Health Care (HI-TRAIN) (Norad Project #: QZA-0484) project. This project, currently in its second of five years, has the following specific aims:

- Provide **post-graduate (Masters and PhD) level training in Health Informatics and research**. Targeted groups

Table 1. **Matrix of capacity-building strategies, likelihood of sustainability and research focus**

Entity targeted	Approach to capacity building			
	Graduate or post-graduate training	Learning by doing	Institutional partnerships between developing and developed countries	Centres of excellence
Individual[a]	+++	+	++	+
Institution	+++	++	+++	+++
Network	++	++	+++	++
National level	+	++	++	+++
Supranational level		++	+++	++
Financial investment[b]	++	+	+++	+++
Research focus →	Research skills ▫▫ → Programme, policy, systems			
Likelihood of sustainability[c] →	+ ——— +++			

[a] + indicates the entity is targeted sometimes; ++ it is targeted moderately often; +++ it is frequently targeted.
[b] Plus signs in this row indicate the extent of financial investment needed by national health research systems or funding agencies: + for low; ++ for medium; +++ for high.
[c] Plus signs in this row indicate the likelihood of sustainability of various approaches: + for fair; +++ for strong.

include health professionals and individuals with computer science background. Personnel trained in HI-TRAIN become core HI faculty at Moi University.

- Increase **number of women and marginalized populations in faculty-level training in Health Informatics and research**.
- Improve the **quality and quantity of Health Informatics research** conducted primarily by researchers based in the LMIC countries in collaboration Northern partners (*see Institutional Partnerships*).
- Provide **model curricula, educational programs and approaches for faculty-level HI training** that can be emulated by regional higher education institutions.

The two-year Masters in Health Informatics program is approved as a university program, and will train an average of 20 candidates per class. In total, 11 Masters candidates will get full scholarship support under HI-TRAIN program to pursue this degree. In return, these individuals will commit to serving as HI personnel and staff within the University. **Table 2** outlines the approved courses for the Masters program. Faculty for this program will come from various departments within Moi University, and from two other partner institutions for HI-TRAIN program, namely Makerere University, Uganda and University of Bergen (UiB), Norway.

Table 2– Approved Masters in Health Informatics Courses

Course Code	Course Title	Contact Hours	Credit Units
HII 811	Introduction to Healthcare and Health Systems	120	3
HII 812	Principles of Public Health	120	3
HII 813	Introduction to Information Technology	120	3
HII 814	Information System Development	120	3
	Total Introductory Courses (2 courses only)	**240**	**6**
MME 801	Innovative Medical Education	40	1
PHM 801	Principles of Management and Health Systems	120	3
HIC 811	Foundations of Health Informatics	120	3
MMR 801	Research Methods	120	3
	Total Core Courses	**400**	**10**
	Total Courses Semester 1	**640**	**16**
PBS 801	Biostatistics	120	3
HIC 822	Health Information Systems	120	3
HIC 823	Health Information Standards and Terminology	80	2
HIC 824	Clinical Decision Support, Ontologies and Workflow	120	3
HIC 825	Scientific Writing and Grantsmanship	80	2
MMP 999	Research Project Design and Implementation (Thesis)	120	3
	Total Courses Semester 2	**640**	**16**
HIC 831	Enterprise Architecture for Health Information Systems	80	2
HIC 832	Law and Governance in Health Informatics	80	2
HIE 831	Public Health Informatics	120	3
HIE 832	Security in Health Systems	120	3
HIE 833	Advanced Programming	120	3
HIE 834	Health Analytics	120	3
HIE 835	Clinical Informatics	120	3
PEC 801	Epidemiology	120	3
	Total Elective Courses (2 courses only)	**240**	**6**
	Total Courses Semester 3	**400**	**10**
HIP 841	Health Informatics Practicum	240	6
MMR 899	Thesis		9
	Total Courses Semester 4		**15**
	GRAND TOTAL OF ALL COURSES TAKEN		**57**

The HI-TRAIN program is also supporting four faculty members to receive full scholarship to undertake PhD training in HI at UiB in Norway, with research conducted within Moi University setting. As part of the scholarship, the faculty will have a binding agreement to serve within the HI programs at Moi to assure sustainability. Moi University has over the last several years also partnered with Regenstrief Institute, Inc. and Indiana University to also offer formal HI fellowship training at Regenstrief Institute. Individuals with fellowship training are now back to oversee new HI programs and research at home. With nine faculty members with fellowship or PhD training in HI, Moi University plans have its own locally run PhD program by 2016. The curriculum development and approval processes for this program start in early 2015.

While the Masters, Fellowship and PhD programs all aim to improve institutional capacity in HI, Moi University has also paid attention to improving capacity in direct response to government and implementation partner needs. Moi University is home to the Regional East African Center for Health Informatics (*REACH-Informatics*). This is project funded by Fogarty International Center as part of the Informatics Training for Global Health (ITGH) Program[11]. Since 2010, REACH-Informatics has offered 25 short courses to 469 participants. Courses have included data management, EHRs developer and implementer training, forms & concept dictionary training, and research database training.

Learning by Doing

Moi University's HI work has been grounded by the need to address real world challenges within the Kenyan setting. The University forms part of a larger care partnership, the Academic Model Providing Access To Healthcare (AMPATH)[12]. AMPATH is a collaborative between Moi University School of Medicine, Moi Teaching and Referral Hospital, and the AMPATH Consortium - a consortium of North American Academic Medical Centers led by Indiana University. Initiated in 2001, AMPATH has used a systems-based approach that closely links clinical care, prevention, research, and training. The program provides comprehensive care and treatment services through 544 MOH facilities and serves a catchment area of three million people.

From its inception, AMPATH recognized the need for comprehensive Health Information Systems to assure optimal care for the people it serves. AMPATH was the first implementation site for the open source electronic record system, OpenMRS, with the implementation called the AMPATH Medical Record System (AMRS)[13]. Success of AMRS helped to convince the Kenyan MOH to roll out OpenMRS in over 300 MOH clinics. Beyond eHealth, AMPATH has used mobile technology for: (a) data collection as part of a home-based counseling and testing for HIV program that has reached over a million individuals; (b) computerized clinical decision support for HIV and chronic disease management; (c) mobile store-and-forward teleconsulttion; and (d) mobile learning and counseling platforms, among others[14-17]. The AMPATH partnership with Moi University offers a real-world laboratory in HI, where trainees can work to complement foundational aspects of HI training. As an innovative arm of the MOH, AMPATH HI innovations largely inform the national conversations in HI.

Institutional Partnerships

A core component to Moi University's HI capacity building initiative is the formation of strong strategic partnerships. These partnerships include:

- *South-South Institutional Partnerships in HI*: Moi University has natural strength in applied Clinical Informatics. The Institution however recognizes that a big part of HI includes foundations of health informatics with a significant computer science component – competencies ranging from algorithms, visualization, analytics, natural language processing, security, to artificial intelligence. To this effect, Moi University has partnered with the School of Computing and Informatics Technology (CIT) of another

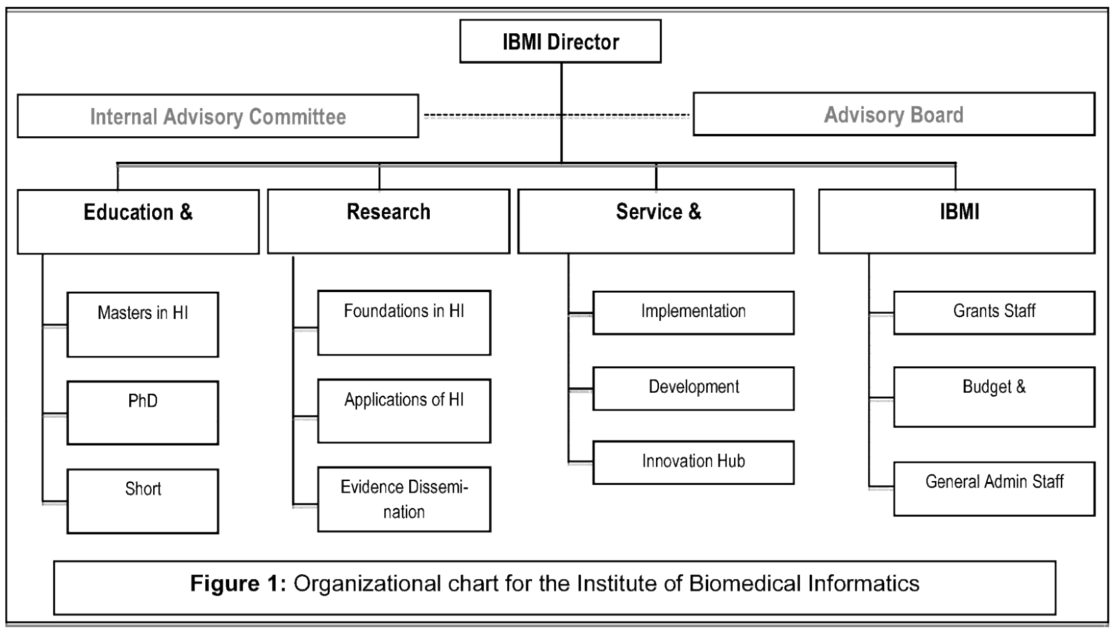


Figure 1: Organizational chart for the Institute of Biomedical Informatics

regional university, Makerere University (Uganda) for its HI initiatives. CIT has 22 PhD holders on the academic staff, with fully fledged academic departments in Information Systems, Computer Science, Information Technology and Software Engineering. Makerere is a member of the HI-TRAIN program, and will be sending students for Masters training in HI at Moi University, and making available faculty to teach in the HI degree courses.

- *North-South Institutional Partnerships in HI:* Over the years, Moi University has partnered with Regenstrief Institute, Inc. (RI) and Indiana University to advance global HI. Among the core achievements of this partnership is the development and implementation of OpenMRS. Moi University has also partnered with University of Bergen (UiB), with a particular emphasis on PhD-training of new emerging HI faculty. All Moi University candidates pursuing PhD-training in HI at UiB will have a Bergen University faculty supervisor in addition to a Moi University supervisor. The HI partnership involves the Center for International Health and Department of Informatics at UiB.
- *Public-Private & Public-Public Partnerships:* As a public institution, Moi University has formed strategic partnerships with organizations in both the public and private sector to ensure sustainability. Partnerships include work with ITECH-Kenya to train implementers on KenyaEMR, which is being rolled out broadly with Kenya. The University is also partnering with Jacaranda Health, a social enterprise for health, on mobile and e-learning platforms for health. Partnerships with counties are in the inception phases, with the goal of training county personnel in HI.

Centers of Excellence

It is well recognized that building centers of excellence offers the best chance for sustainability and long standing capacity building efforts[6]. Centers provide academic homes for highly trained faculty in HI and allow for consolidation of numerous approaches to capacity building. To this end, Moi University has established an Institute of Biomedical Informatics (IBMI), to serve as a regional HI Center of Excellence. The IBMI at Moi is made up of: (a) *Education & Training Division* – hosting the Masters program, PhD program, and short courses; (b) *Research Division* – serving as the research and dissemination arm; (c) *Service & Extension Division* – providing consulting, development and implementation services to MOH and partners in the region; and (d) *Administration Division* – overseeing all administrative aspects of the IBMI (**Figure 1)**. IBMI is an integral part of the University, with core funding coming from the government of Kenya to assure sustainability. Additional sources for Institute financial sustainability come from student tuition, research grants, and service-based fees.

Discussion

Existing approaches to HI capacity building within resource-limited settings remain largely non-comprehensive. In this paper, we present a case study of a publicly-funded higher education institution in a developing country that has taken a multi-dimensional HI capacity building approach guided by a comprehensive framework. This approach incorporates training, learning by doing, partnerships and a center of excellence. The approach can serve as a model for adoption by other institutions in similar resource-limited settings.

There are multiple reasons why the described capacity-building model needs to be the goal for institutions aiming to improve HI capacity in developing countries. To improve HI capacity, countries need highly trained individuals to serve as teachers and trainers. These individuals should ideally be locally-based, and should have a supportive home-institutional environment within which to work. The work should be grounded by the country HI capacity needs and strategies, and by interaction with real-world healthcare settings to inform innovations and evaluations. In a resource-limited setting, chances of self-sufficiency and sustainability can only happen when strategic partnerships are formed to reduce burden to individual institutions. Ideally, core support should come from the country's ministries of education and health, as this will provide an excellent foundation for sustainability, and for alignment with country's HI systems strategy.

The described capacity building approach has its limitations. To implement all dimensions for capacity-building, a significant amount of financial investment is needed. It is fair

to say that this capacity-building model is not necessarily practical for all institutions. It is one best suited for select institutions tasked with a larger role of leading each country's HI capacity building initiatives. It should also be noted that each component of the model can be used to generate financial resources for sustainability. Tuition will be charged for training and degree programs, service fees can be incorporated into the learning-by-doing experience and into centers of excellence activities. Partnerships can also be strategically used to alleviate the financial burden to the institution.

To actualize the proposed comprehensive approach takes time. It would be unrealistic to expect institutions to implement all dimensions at once or within a short period. At the very least, institutions should have a roadmap that would eventually see the actualization of all dimensions over realistic timelines. Alternatively, multiple institutions with complementary strengths can collaborate to create a comprehensive partnership that has every dimension of the model, without necessarily having all components fall under a single institution. As an example, an institution with strengths in HI training can partner with another institution with a HI Center of Excellence.

Funding organizations should pay attention to the models for capacity building they are supporting. The NORHED capacity building initiative by NORAD is a program that is already looking at capacity building initiatives at very comprehensive level. In addition to encouraging all dimensions, the NORHED program also emphasizes gender equity, as females are often underpresented in the Sciences. Other programs that can have impact in this area include ITGH and the Rockefeller's Digital jobs initiative. The goal should be sustainable HI capacity initiatives beyond a grant period.

Conclusion

It is possible to build comprehensive, locally-run HI capacity building initiatives that are sustainable, in line with country capacity building priorities, and which offer high quality and accredited HI training.

Acknowledgments

This work was supported in part by NORHED program (Norad: Project # QZA-0484) and the REACH-Informatics Program (Fogarty Institute: Grant # D43 TW008447-01S1). Special thanks to the Vice Chancellor, Moi University (Prof. Richard K. Mibei); Dean, School of Information Sciences (Prof. Justus Wamukoya); and core HI partner institutions, namely Makerere University, Regenstrief Institute, Inc., and University of Bergen.

References

[1] Williams F, Boren SA. The role of the electronic medical record (EMR) in care delivery development in developing countries: a systematic review. Informatics in primary care. 2008;16(2):139-45.

[2] Hersh W, Williamson J. Educating 10,000 informaticians by 2010: the AMIA 10x10 program. International journal of medical informatics. 2007;76(5-6):377-82.

[3] The Informatics Training for Global Health Program of the Fogarty International Center. Available at http://www.fic.nih.gov/programs/training_grants/itgh/. Last accessed on Dec-15-2014.

[4] Curioso WH, Fuller S, Garcia PJ, Holmes KK, Kimball AM. Ten years of international collaboration in biomedical informatics and beyond: the AMAUTA program in Peru. Journal of the American Medical Informatics Association : JAMIA. 2010;17(4):477-80.

[5] Kimball AM, Curioso WH, Arima Y, Fuller S, Garcia PJ, Segovia-Juarez J, et al. Developing capacity in health informatics in a resource poor setting: lessons from Peru. Human resources for health. 2009;7:80.

[6] Lansang MA, Dennis R. Building capacity in health research in the developing world. Bulletin of the World Health Organization. 2004;82(10):764-70.

[7] Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. Journal of the American Medical Informatics Association : JAMIA. 2012;19(6):931-8.

[8] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods of information in medicine. 2010;49(2):105-20.

[9] Hasman A, Mantas J. IMIA Accreditation of Health Informatics Programs. Healthcare informatics research. 2013;19(3):154-61.

[10] Moi University, Eldoret, Kenya. https://http://www.mu.ac.ke/. Last Accessed 12-15-2014.

[11] Informatics Training for Global Health (ITGH) Program of the Fogarty International Center. http://www.fic.nih.gov/programs/pages/informatics.aspx. Last accessed on 12-15-2014. .

[12] Inui TS, Nyandiko WM, Kimaiyo SN, Frankel RM, Muriuki T, Mamlin JJ, et al. AMPATH: living proof that no one has to die from HIV. Journal of general internal medicine. 2007;22(12):1745-50.

[13] Tierney WM, Rotich JK, Hannan TJ, Siika AM, Biondich PG, Mamlin BW, et al. The AMPATH medical record system: creating, implementing, and sustaining an electronic medical record system to support HIV/AIDS care in western Kenya. Studies in health technology and informatics. 2007;129(Pt 1):372-6.

[14] Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. Journal of the American Medical Informatics Association : JAMIA. 2012;19(4):655-9.

[15] Were MC, Shen C, Tierney WM, Mamlin JJ, Biondich PG, Li X, et al. Evaluation of computer-generated reminders to improve CD4 laboratory monitoring in sub-Saharan Africa: a prospective comparative study. Journal of the American Medical Informatics Association : JAMIA. 2011;18(2):150-5.

[16] Were MC, Nyandiko WM, Huang KT, Slaven JE, Shen C, Tierney WM, et al. Computer-generated reminders and quality of pediatric HIV care in a resource-limited setting. Pediatrics. 2013;131(3):e789-96.

[17] Anokwa Y, Ribeka N, Parikh T, Borriello G, Were MC. Design of a Phone-Based Clinical Decision Support System for Resource-Limited Settings. Information and Communication Technologies and Development ACM International Conference Proceeding Series. 2012:13-24.

Address for correspondence

Martin C. Were, MD MS
Regenstrief Institute Inc., & Indiana University School of Medicine
410 West 10th Street, Suite 2000, Indianapolis, IN 46202

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-525

Health Informatics: Developing a Masters Programme in Rwanda based on the IMIA Educational Recommendations and the IMIA Knowledge Base

Graham Wright[a], Frank Verbeke[bc], Marc Nyssen[bc], Helen Betts[a]

[a] *CHIRAD, University of Fort Hare, East London, South Africa*
[b] *Department of Biostatistics and Medical Informatics, Vrije Universiteit, Brussel, Brussels, Belgium*
[c] *Regional e-Health Center of Excellence, University of Rwanda, Kigali, Rwanda*

Abstract

Since 2011, the Regional e-Health Center of Excellence in Rwanda (REHCE) has run an MSc in Health Informatics programme (MSc HI). A programme review was commissioned in February 2014 after 2 cohorts of students completed the post-graduate certificate and diploma courses and most students had started preparatory activity for their master dissertation.

The review developed a method for mapping course content on health informatics competences and knowledge units. Also the review identified and measured knowledge gaps and content redundancy. Using this method, we analyzed regulatory and programme documents combined with stakeholder interviews, and demonstrated that the existing MSc HI curriculum did not completely address the needs of the Rwandan health sector. Teaching strategies did not always match students' expectations. Based on a detailed Rwandan health informatics needs assessment, International Medical Informatics Association (IMIA)'s Recommendations on Education in Biomedical and Health Informatics and the IMIA Health Informatics Knowledge Base, a new curriculum was developed and provided a better competences match for the specifics of healthcare in the Central African region. The new approved curriculum will be implemented in the 2014/2015 academic year and options for regional extension of the programme to Eastern DRC (Bukavu) and Burundi (Bujumbura) are being investigated.

Keywords:

Education; Medical Graduate; Teaching; Curriculum; Knowledge.

Introduction

The International Medical Informatics Association (IMIA) is the international organization for health and biomedical informatics. IMIA acts as a bridging organisation, bringing together the constituent regional and country health informatics associations and their members. The IMIA General Assembly has approved a number of endorsed documents:

1. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics; First Revision (2010)
2. IMIA Knowledge Base (2011)

The Recommendations of the IMIA on Education (1)

Rather than reproduce all the recommendations, the first 6 illustrate the nature of the recommendations:

Biomedical and Health Informatics Core Knowledge and Skills

1.1 **Evolution of informatics** as a discipline and as a profession

1.2 **Need for systematic information processing** in health care, benefits and constraints of information technology in health care

1.3 Efficient and responsible **use of information processing tools**, to support health care professionals' practice and their decision making

1.4 **Use of personal application software** for documentation, personal communication including Internet access, for publication and basic statistics

1.5 **Information literacy**: library classification and systematic health related terminologies and their coding, literature retrieval methods, research methods and research paradigms

1.6 Characteristics, functionalities and examples of **information systems in health care** (e.g. clinical information systems, primary care information systems, etc.)

The IMIA Knowledge Base (2-4)

In 2011, IMIA also published a knowledge base structuring the knowledge of the Health Informatics domain into 14 topics (listed below). The base was subdivided into 245 knowledge units. This unit list served in the review process as a tool for measuring the coverage of Health Informatics knowledge by the MSc HI curriculum:

1. Computer Science for Health Informatics (ICT for Health)
2. Health & Social Care Processes
3. Health (care) Records
4. Health and Social Care Industry
5. Health Informatics Standards
6. Knowledge Domains and Knowledge Discovery
7. Legal and Ethical
8. People in Organisations
9. Politics and Policy
10. Technologies for Health
11. Terminology, Classification and Grouping
12. Uses of Clinical Information
13. Using Informatics to Support Clinical Healthcare Governance
14. Computer Systems Applications in Health (Toolkit)

Figure 1 – Section of the IMIA Knowledge Base

In February 2014 in Kigali with a team of 3 international reviewers, the University of Rwanda, College of Medicine and Health Sciences (CMHS) commissioned a curriculum review of the MSc in Health Informatics (MSc HI) programme. The team used an Action Research method to review and rewrite the curriculum based on the two IMIA endorsed documents above.

The review activities consisted of:

- A study of national regulations and documents relevant for the MSc HI programme.
- A study of the MSc HI programme documents.
- A study of international standards, recommendations and literature on health informatics postgraduate education programmes (6,7) especially those that had been endorsed by the IMIA.
- A mapping of the existing MSc HI curriculum content on to international standards and regulations.
- A field study of needs and competences in Rwanda in the area of health informatics (5), mainly through a number of interviews with relevant stakeholders, including Ministry of Health (MoH) staff, NGOs, public and private health facilities, education & training institutions, health software development companies and research organisations.
- An evaluation of how students, teaching staff and CMHS faculty perceived the existing MSc HI curriculum.

Information obtained from the MoH clearly showed a growing need for people with combined skills covering Information Technologies (IT), Information Systems (IS), Information Management (IM) and Health System (HS) knowledge. Generally speaking, there is a shift from IT and IS towards IM and HS needs: internal MoH software and systems development activities are progressively phasing out in favour of private sector outsourcing. Consequently, the MoH requires new management skills related to contracting, people management, and project management. Health facilities identified competences related to health care research (mainly for university teaching hospitals), data mining, and statistical and epidemiological reporting based on Electronic Health Record (EHR) content. Public and private hospitals described the urgent need for Chief Information Officer (CIO) profiles that again must be capable of combining IT, IS and IM knowledge and skills.

Based on competence needs and employment perspectives, the following human resource profiles can be identified for MSc HI graduates:

- Health Information Managers (mainly IM and HS knowledge and skills needed in public sector)
- Clinicians proficient in IT, IS and IM (health facilities)
- Information System Implementers and System Integrators (e.g. for EMR-HMIS integration)
- Information System Evaluators (Health Informatics Interventions impact measurement and Information Systems quality evaluation)
- Education staff (Academic staff and Information Systems training staff / functional experts)
- Researchers (Health Informatics domain and other Health Care domains)

All of the competences identified by the Rwandan stakeholders were covered by the IMIA recommendations on Education.

Methods

The IMIA recommendations were mapped to the IMIA Knowledge Base and the learning outcomes produced would underpin the identified human resource profiles. This mapping then lead the team to develop an SQL database.

IMIA competences - knowledge base mapping

In the first exercise, 40 competences identified in the "IMIA recommendations for health informatics education" were mapped on 0 or more knowledge units in the IMIA knowledge base. This resulted in a total of 400 mappings, clearly demonstrating the IMIA Knowledge Base units synergy with the IMIA competences.

IMIA competences - existing programme mapping

Each module title of the existing MSc HI programme was then mapped onto a number of IMIA competences (replacing the original programme objectives). This showed that a total set of 38 IMIA competences (out of 40) were covered by the existing programme. The existing programme modules did not have consistency within or between modules in terms of content. There were gaps in the content and poorly distributed weightings of the content. However, the module titles and objectives gave a good starting point to automatically populate these modules using the IMIA recommendations and the IMIA Knowledge Base. The computer based exercises took 5 days using a three dimensional SQL Querying method devised by Dr. Verbeke.

IMIA knowledge - existing curriculum mapping

Based on the mapping between existing MSc HI module titles and IMIA competences combined with the previously performed mapping between IMIA competences and IMIA knowledge units, an automatic mapping between existing MSc HI modules and IMIA knowledge units was generated. This demonstrated that 219 (out of 245) knowledge units were covered by the existing titles and structure.

Development of a reference curriculum

The automatically generated mapping table between existing curriculum modules and IMIA knowledge units was then cleaned for redundancy (assigning each knowledge unit to only one single module) and for completeness (modifying module content in order to cover previously missed knowledge units). This provided a set of modules with 100% knowledge coverage whilst avoiding any redundant or repetative teaching.

In a final phase, a number of modules were (partially) merged in order to come to an equally distributed module weight of 10

credits for each module, with 6 modules in the post-graduate certificate year and 6 modules in the post-graduate diploma year.

Results

An Example of Module Aims, Outcomes and Content Introduction to Health Informatics

Brief description of aims and content

This module introduces the discipline of health informatics and the details of the programme. The module covers a number of basic concepts in health informatics and ensures that students from different backgrounds have the opportunity to share perceptions of health and medical informatics.

Learning outcomes (competences) from IMIA Recommendations

1.1 Evolution of informatics as a discipline and as a profession

1.4 Use of personal application software for documentation, personal communication including

Internet access, for publication and basic statistics

3.1 Basic informatics terminology like data, information, knowledge, hardware, software, computer, networks, information systems, information systems management

3.2 Ability to use personal computers, text processing and spreadsheet software, easy-to-use database management systems

3.3 Ability to communicate electronically, including electronic data exchange, with other health care professionals, Internet /intranet use

3.4 Methods of practical informatics/computer science, especially on programming languages, software engineering, data structures, database management systems, information and system modelling tools, information systems theory and practice, knowledge engineering, (concept) representation and acquisition, software architectures

3.6 Methods of technical informatics/computer science, e.g., network architectures and topologies, telecommunications, wireless technology, virtual reality, multimedia

Indicative content (units) from IMIA Knowledge Base

- Access to information
- Computer literacy (ECDL)
- Computer systems
- Computing methodologies
- Data management and storage
- Databases
- Demystify IT for users
- Explains health informatics
- History of methods of gathering information in the clinical workplace
- Human Computer Interaction (HCI) principles
- Information sources
- Information Storage and Retrieval
- Internet, intranets and associative technologies
- Networking
- Web technologies
- Wireless technology
- Prototype system for a department

The curriculum content was also weighted according to the recommended student workload in European Credit Transfer and Accumulation System (ECTS) credits for the knowledge and skill areas of a two year biomedical and health informatics master (BMHI) programmme.

Table 1 - European Credit Transfer and Accumulation System (ECTS) credits

Knowledge/Skill Area	Credits
1. BMHI Core Knowledge And Skills	80
2. Medicine, Health And Biosciences, Health System Organisation	20
3. Informatics/Computer Science, Mathematics, Biometry	40
Total ECTS credits	120

The teaching components of the programme total 120 credits. These components are divided between the Certificate and Diploma levels. However the number of ECTS recommended by the IMIA team are 140; therefore, 20 of the credits will be included in the Research Thesis.

Post-graduate certificate

The post-graduate certificate programme essentially focuses on IT and IS after an introduction to the Health Informatics domain. The development of practical skills and essential knowledge of the Rwandan health informatics context are core components of the programme.

Introduction to Health Informatics: includes an introduction to the discipline of Health Informatics and the scope of the programme. The course covers a number of basic concepts in Health Informatics and ensures that the students have the opportunity to share perceptions of Health and Medical Informatics (8).

Healthcare Management and Organisation: covers the concepts of Healthcare Management in relation to Health Informatics. The Health Informatics concepts underpinning management, organisational culture and socio-technical aspects are explored (8).

Knowledge Management in Healthcare Delivery: covers Health Informatics topics, in particular the processing of Data into Information, and then into Knowledge. The use of this knowledge is explored in Decision Making and Education.

Management Information Systems: explores the connection between information systems (IS) and business performance. It also explores the issues of security, transparency, traceability and Return on Investment.

Electronic Health Records Management and Hospital & Health Information Systems: assists students to understand the complexities of managing individual and community based health information. Students will be able to understand different approaches of health record modeling; purpose based structuring of health information and the principles of integrated health information management (9).

Health Informatics Applications including PACs, MIT, Telemedicine and mobile technologies: explores the concepts of ubiquitous computing technologies to provide positive support in the population. It explores the use of telemedicine and e-Health solutions for medical support. This is underpinned by an exploration of wireless technologies.

Post-graduate diploma

The post-graduate diploma year is designed to add more IM knowledge and skills, leading the student to a broader comprehension of the complete health informatics domain and its place in the health system ("the big picture"). It also adds

essential project management skills and elements of critical thinking enabling well-considered evaluation of different options. Research skills and scientific reasoning are formally included in the programme.

The following modules are part of the diploma programme:

Software Based Clinical Decision-making and Support Systems: explores the concepts of information to support health professionals in their decision-making and development of therapeutic strategies. The module covers clinical pathways and guidelines and biomedical modeling and simulation as well as the principles of data representation and cognitive aspects of information processing.

Public Health Informatics includes Patient Information Kiosks, Websites and Public Health Systems for Epidemiology, Epidemic Control and GIS: explores the concepts of data representation and information analysis. The module includes information tools to support education for the public and health professionals, epidemiology and public health.

Social-Cultural, Legal and Economic Impact of Health Informatics: explores ethical, legal, and social issues arising in the use of computer-based technology and information systems in the delivery of health care. The module also includes health informatics ethics and regulatory frameworks.

Project Management and IT Introduction in Health Care Delivery, Case Studies in Health Informatics: students learn the principles of project management and using information management within projects. Through exploration of health information projects, students gain a real-world understanding of how to manage biomedical informatics projects taking into account the socio-technical aspects of implementation.

Research, Monitoring and Evaluation in Health Informatics: explores an introduction to research in the areas of health informatics and covers a wide range of methods and techniques. The topics range from epidemiology data and analysis and mapping to measures of patient outcomes following medical interventions. Evaluation and monitoring of programs of care and clinical audit are also included (9).

E-health Enterprise Architecture: explores the details of healthcare information technology (HIT) interoperability and standards. The evolution of technology in healthcare, along with the impact on clinical information systems, is studied. The benefits of integrating healthcare information systems are investigated (9), as are the challenges of integrating systems across disparate organisations, healthcare disciplines, and technologies. The value proposition of a standards-based approach to integration is presented (10).

Discussion

This curriculum has been validated and plans to start in January 2015. The curriculum is the first in Africa to be based on the IMIA recommendations.

The new MSc HI programme has been tailored to the Central African context (8) and options for regional extension of the program to Eastern DRC (Bukavu) (10) and Burundi (Bujumbura) are being investigated.

Conclusion

The 3-dimensional mapping of IMIA competences, IMIA knowledge base units and curriculum module titles provided a useful method for validating biomedical and health informatics domain coverage of a competence based MSc HI curriculum in Rwanda. The same method could be applied to other MSc HI programmes seeking compliance with IMIA biomedical and health informatics education standards. Unfortunately there is insufficient space to include details of the SQL queries in this article. The authors intend to publish this in a technical paper.

Many of the content gaps found by our analysis might have been revealed earlier had more robust evaluation and monitoring processes been in place. In addition to the work discussed in this paper, the review also explored the teaching and assessment strategies employed. The programme content gaps and teaching strategy issues discovered during the curriculum review demonstrated the importance of periodical external curriculum reviews.

References

[1] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, et al. IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. First revision. Methods Inf Med 2010: 49(2): 105-20.

[2] Wright G. The Development of the IMIA Knowledge Base. SA Journal of Information Management 2011: 13(1): 5.

[3] Wright G, Betts H, and Murray P. Defining the Discipline of Health Informatics: An illustration of triangulation via mixed methods in a study. Advancing Knowledge for empowerment through innovations in research; ICC, East London, South Africa: Walter Sisulu University, 2010.

[4] Murray PJ, Wright G, and Lorenzi NM. Developing the Knowledge Base for Health Informatics: From the Otley Think-tank to the IMIA Strategic Plan. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems 2007: 2263.

[5] Verbeke F. Contributions to the Development of Health Informatics to address e-Health Challenges in Rwanda, Rwanda Medical Journal 2012: 69(4).

[6] Berner E, Informatics Education in Healthcare: Lessons Learned, Springer: London, 2014.

[7] Hersch W, Margolis A, Quirós F, and Otero P. Building a health informatics workforce in developing countries. Health Affairs 2010: 29: 274-7.

[8] Luna D, Otero C, and Marcelo A. Health Informatics in Developing Countries: Systematic Review of Reviews. Contribution of the IMIA Working Group Health Informatics for Development. Yearbook of medical informatics 2012: 8: 28-33.

[9] Verbeke F, Karara G, and Nyssen M. Evaluating the Impact of ICT-tools on Health Care Delivery in Sub-Saharan Hospitals. Stud Health Technol Inform 2013: August 22.

[10] Verbeke F, Shamashanga P, Amisi C, and Karara G. The National Plan for Health Informatics Development in DRC, CEDESURK, Kinshasa, 2014.

Address for correspondence

Professor Graham Wright, CHIRAD, School of Health Sciences, University of Fort Hare, Box 294, Bathurst, South Africa. Prof.Graham.Wright@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-529

Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Term Coverage

David T. Marc[a], Rui Zhang[a,b], James Beattie[c], Laël C. Gatewood[a], Saif S. Khairat[a,d]

[a] *Institute for Health Informatics - University of Minnesota- Twin Cities, Minneapolis, MN*
[b] *Department of Surgery - University of Minnesota- Twin Cities, Minneapolis, MN*
[c] *Bio-Medical Library- University of Minnesota- Twin Cities, Minneapolis, MN*
[d] *Carolina Health Informatics Program- University of North Carolina, Chapel Hill, NC*

Abstract

As part of the Open Government Initiative, the United States federal government published datasets to increase collaboration, transparency, consumer participation, and research, and are available online at HealthData.gov. Currently, HealthData.gov does not adequately support the accessibility goal of the Open Government Initiative due to issues of retrieving relevant data because of inadequately cataloguing and lack of indexing with a standardized terminology. Given the commonalities between the HealthData.gov and MEDLINE metadata, Medical Subject Headings (MeSH) may offer an indexing solution, but there needs to be a formal evaluation of the efficacy of MeSH for covering the dataset concepts. The purpose of this study was to determine if MeSH adequately covers the HealthData.gov concepts. The noun and noun phrases from the HealthData.gov metadata were extracted and mapped to MeSH using MetaMap. The frequency of no exact, partical and no matches with MeSH terms were determined. The results of this study revealed that about 70% of the HealthData.gov concepts partially or exactly matched MeSH terms. Therefore, MeSH may be a favorable terminology for indexing the HealthData.gov datasets.

Keywords:

MeSH; Data; Standards; Indexing; Data Storage and Retrieval.

Introduction

The United States federal government collects and stores massive amounts of data, yet that data has largely been inaccessible to the American people [1]. In an effort to change the culture regarding governmental transparency and information sharing, then Presidential-candidate Barack Obama pledged to create a transparent and connected democracy [2]. On January 21, 2009, the day President Obama was sworn into office, the "Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government" was signed [3]. This Memorandum outlined the Obama Administration's "Open Government Directive" that detailed actionable steps and deadlines that federal agencies were required to take to promote the principles of transparency, participation and collaboration. This directive initiated the Data.gov website [4] for publishing and disseminating federally-sponsored datasets. The memorandum specified that each federal agency must identify and publish in an open format at least three high-value data sets and register those datasets via Data.gov [5].

High value information was defined as "information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation" [5]. In the U.S. National Action Plan, a policy was instated regarding "smart disclosure," or releasing high quality, complex information in standardized, machine-readable formats that can improve decisions and aid scientific research. The action plan noted the government's intent to develop Federal guidelines on scientific data [6]. Federal agencies were encouraged to release a variety of "smart disclosure" data that was accessible, machine-readable, standardized, timely, adaptive and innovative to markets, in an interoperable, and deidentifed format [7].

Publicly Available Data

Data.gov is an online data repository that includes tools and resources, communities, and education on how to utilize the site and its data. A community portal, HealthData.gov [8], was specifically developed to support the dissemination of healthcare data including Medicare and Medicaid cost reports, public health registries, medication treatments, and county-level determinants-of-health [8]. The overall impact of the Open Government Initiative on healthcare data resulted in the release of almost 2000 high value data sets on HealthData.gov.

In both Data.gov and HealthData.gov, users have the ability to filter datasets by subject, description, publisher, keywords, and other relevant meta-data (Table 1). Although the dissemination of publically available data is a step forward in promoting governmental transparency, there exist significant limitations in actually acquiring the datasets. Specifically, the limitations in the annotation principles adopted by HealthData.gov may result in suboptimal results from data searches.

Data.gov and HealthData.gov adopted a data management system constructed by Comprehensive Knowledge Archive Network (CKAN), an open-source data portal software [9]. The CKAN search application programming interface (API) is based on keyword searches for querying and accessing data. The searches are based on metadata that is supplied by the data publisher. To the best of our knowledge, the performance of the CKAN search API has never been formally evaluated.

Public Data Challenges

There are obvious reasons why the federal government is making high quality healthcare data available for research and use by the public, including benefits related to the goals of

collaboration, transparency, participation, and research. The overarching benefit of these data is the potential to improve decisions based upon knowledge [7]. However, the impact of the open government initiative on supporting research has been minimal. A primary reason why the open government initiative is failing to support research is due in part to the fact that transparency and access standards are inconsistently applied within federal communication policymaking [10]. Thus, the lack of consistency may be inhibiting citizen engagement in utilizing these resources for scholarly purposes [11]. In fact, studies have shown that although the government is adopting technology to increase transparency, these tools have been ineffective in engaging citizen participation [12-15].

Table 1 - Primary Meta-data Fields that are required for all Datasets in HealthData.gov

Field	Description
id	The unique identifier for the dataset
title	The title name for the dataset
notes	The description of the dataset
author	The name of the federal agency that submitted the dataset
tags	Tags associated with the dataset

Although the websites Data.gov and HealthData.gov have the potential to support governmental transparency, the success of these sites rests upon how well they can accommodate consumer accessibility [16]. Robinson and colleagues [17] discussed how full compliance and coordination with government-wide policy requirements present challenges for federal websites. Chang and Kannan [18] and Dawes and Helbig [14] suggested that more research is needed to explore the needs of citizens in order to enhance data tools. Similarly, others have suggested that there needs to be a more integrated approach to investigate data accessibility and dissemination tools and inventory resources and solutions to any problems [15, 18].

Information accessibility remains a challenge for citizens regardless of the growth of Data.gov and HealthData.gov. Research is needed to investigate practices that can accommodate consumers in accessing and using publically available federal data. Studies have suggested that the government should allow third party, private entities to play larger roles in the delivery of data and information [15, 17]. Feasibly, other federal efforts can be leveraged to support accessibility and usability of publicly available data. In Europe, several initiatives have begun to support the sharing of data. OpenPHACTS is one example of a partnership between public and private stakeholders from academia, pharmaceutical companies, and other enterprises that is funded through the Innovative Medicine Initiative to support common standards and sharing of data for more efficient drug development and patient treatment [19]. The European Medical Information Framework (EMIF) is another initiative that aims to create an environment for the re-use of existing health data. The EMIF Platform allows researchers to browse available patient-level data in Europe to support research. A data repository known as tranSMART is currently being developed to provide a schema for a variety of data types [20].

As Data.gov and HealthData.gov expand, the ability for end-users to find relevant data sources will become paramount for the sustainability of the websites and the Open Government Initiative. Particularly, the data extraction methods that were adopted in the construction of Data.gov and HealthData.gov need to be evaluated so that the most relevant data is queried while also avoiding irrelevant data. Additionally, proactive indexing of datasets should be systematic, structured, and not too labor-intensive to allow for long-term sustainability and usability of publically available data.

Medical Subject Headings (MeSH)

In the 1960s, the U.S. National Library of Medicine (NLM) dealt with information retrieval issues around biomedical literature due to poor indexing and search mechanisms for articles in MEDLINE. Due to problems in retrieving relevant medical literature, NLM produced MeSH. The goal of MeSH is "to provide a reproducible partition of concepts relevant to biomedicine for purposes of organization of medical knowledge and information" [21].

MeSH is one of several biomedical vocabularies that are available in the Unified Medical Language System (UMLS) Metathesaurus. The UMLS Metathesaurus is a large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health related concepts, their various names, and the relationships among them [22].

Tools were developed that utilize the UMLS Metathesaurus. MetaMap is a natural language processing (NLP) tool that was developed to map biomedical text to the UMLS Metathesaurus [23]. Input text is processed by MetaMap using lexical/syntactic analyses, followed by variant generation, candidate identification, mapping construction, and word sense disambiguation. There are processing options that allow users to limit the UMLS Metathesaurus to a specific vocabulary, such as MeSH.

MeSH is based off of a concept-oriented framework. The framework has a three-level structure: descriptor class, concept, and term. A concept is a common idea or meaning expressed by synonymous words or terms. A concept is a term for each of these synonymous terms. A term is a way of describing the concept. The concept will often have a preferred term, which has been adopted as the name of the concept in the MeSH framework. The descriptor is a class of one or more related concepts where the name of the descriptor is the preferred concept.

The main headings and entry terms are the names of the concepts in a descriptor class. In MEDLINE, the descriptor name is often attached to the citation record. An entry term may be the descriptor name or a concept in that descriptor class. The role of entry terms is to provide a guide for choosing proper concepts. Therefore, searchers most often use entry terms for determining the most appropriate concept class.

Interestingly, there are very few studies evaluating the performance of the MeSH framework for retrieving resources other than medical literature [24, 25]. Of particular interest is the possible influence of the MeSH vocabulary in indexing and retrieving publically available healthcare data that is published under the Data.gov community site, HealthData.gov.

Indexing Public Data with MeSH

The metadata that is available in CKAN offers similar information that is available within biomedical literature published on MEDLINE, including a title, author, description/abstract. CKAN recommends that publishers provide specific metadata to describe the datasets (Table 1). However, CKAN does not currently support a standardized vocabulary for indexing purposes. Tags, also known as keywords, are a metadata element provided in CKAN that is self-prescribed by the authors of those datasets. In constrast, MEDLINE uses MeSH entry terms for indexing biomedical literature. Currently, MeSH entry terms are assigned to each

article that is accessible through MEDLINE by trained indexers hired by the U.S. NLM. However, efforts for automating the indexing process in MEDLINE have shown promise.

The Medical Text Indexer (MTI) was developed for the sole purpose of offering MeSH concepts by automatically processing the MEDLINE metadata [26]. Research suggested that biomedical literature in MEDLINE was successfully indexed with MTI when only a title and abstract were provided [26]. Interestingly, the auotmated indexing engine, MTI, was shown to have similar accuracy as a human indexer, but with far greater efficiency [26]. Therefore, MTI was shown to be a substitute to human indexing, which supports a more efficient indexing process. The MTI may also prove to be an efficient and accurate tool for indexing datasets that are available through HealthData.gov. However, before an automated indexing engine can be adopted and applied to the HealthData.gov metadata, there is a need to evaluate the alignment of MeSH entry terms to the concepts that are present in the HealthData.gov metadata.

To support the adoption of MeSH for indexing and retrieving healthcare data, a formal evaluation is needed to determine if MeSH is an appropriate vocabulary for data indexing and retrieval. The purpose of the proposed study is to test this thesis. MeSH should only be adopted if the vocabulary offers consistent and adequate coverage of the concepts that are available in HealthData.gov. This study will explore the coverage of terms by parsing out the noun and noun phrases from the HealthData.gov metadata and mapping these terms to the MeSH vocabulary.

Methods

Extracting Nouns and Noun Phrases from HealthData.gov

The complete HealthData.gov metadata catalogue from datasets that were published from May 2012 to May 2014 were queried from the CKAN engine using an HTTP GET script. There was a total of 1,003 datasets published during this time. The HealthData.gov metadata was exported from CKAN in a JSON file format. The *json* Python library was used to parse the JSON file to extract the notes (*i.e.,* the descriptions) from the metadata. The extracted metadata were written to a CSV file.

The metadata CSV file was mapped to the UMLS Metathesaurus by using MetaMap to identify the nouns and noun phrases. The supplied text was parsed into simple noun phrases using the SPECIALIST minimal commitment parser [27]. Variants were generated for each phrase using the SPECIALIST lexicon. Variants consisted of phrase words together with their acronyms, abbreviations, synonyms, derivational variants, and meaningful combinations of these, including inflectional and spelling variants [28]. A candidate set of all Metathesaurus strings containing at least one of the variants was retrieved. Each candidate was evaluated against the input text by mapping the phrase words to the candidate's words and then calculating the strength of the mapping using a linguistically principled evaluation function [29].

To determine the overall frequency of all nouns and noun phrases that occured in the HealthData.gov notes, all UMLS concepts were indexed with MetaMap and the output was selected to reveal noun phrases, nouns, and acronyms/abbreviations. The MedPost/Semantic Knowledge Representation (SKR) part-of-speech tagger broke the original notes for each dataset into sentences, tokenized each sentence and then further tagged tokens as nouns, verbs, prepositions, adjectives, and punctuations. Candidates were identified from the list of variants for each concept. Once the nouns and noun phrases were identified, MetaMap provides UMLS concept candidates with concept unifier identifiers (CUIs), preferred names and semantic types. All of the mapping results were printed and exported in a machine-readable output. The machine-readable output was further processed by extracting nouns from the summary while omitting verbs, prepositions, adjectives, and punctuation using a regular expression function in the R statistical software package. The frequency of nouns, noun phrases, and distinct nouns and noun phrases that were present in the HealthData.gov metadata were then reported.

MeSH Coverage of Nouns

The complete MeSH library was downloaded as an XML file from the U.S. NLM webpage [30]. The XML file was parsed using the Python-based Element Tree XML API to create a Python dictionary of all the MeSH concepts, related concepts, and entry terms. The dictionary was written to a CSV file and uploaded to R to calculate MeSH term coverage.

In order to determine the level of coverage that MeSH offered for indexing the HealthData.gov metadata, the nouns that were extracted from the metadata using MetaMap were mapped to MeSH entry terms. The nouns were mapped to MeSH by developing a query in R to search for partial or exact matches between the nouns extracted using MetaMap and the complete MeSH library. The frequency of partial, exact, and no matches between the MeSH terms and the extracted MetaMap nouns were calculated. If a noun was contained in part of a MeSH entry term, it was considered a partial match. A noun that exactly matched a MeSH entry term was considered an exact match. Any noun that was not part of a MeSH entry term was considered not matching.

Further analysis was conducted by restricting the comparison of MeSH terms to the nouns extracted using MetaMap for those nouns that accounted for 75% of all of the identified nouns. The 75% threshold was determined based on the observation that this accounted for nouns that occurred at least 10 times in the HealthData.gov metadata. This was completed in order to evaluate the efficacy of indexing the data with MeSH while minimizing the effects of nouns that do not occur frequently in the HealthData.gov metadata. The frequencies of partial, exact and no matches were calculated to determine the efficacy of MeSH for indexing the most common occurring nouns in HealthData.gov.

Common MeSH Concepts in HealthData.gov

A final analysis was carried out to determine the most common MeSH concepts that occured in the HealthData.gov metadata. The HealthData.gov metadata CSV file was processed a second time using MetaMap, but restricting the library source to only MeSH. The phrases that were mapped to MeSH were returned along with the MetaMap summary, which was exported to show the associated MeSH candidate concepts that map to the noun phrases. A mapping score was provided that revealed the level of agreement between a noun phrase and the MeSH candidate concept. A mapping score of 1000 indicated an exact match. In addition to the MeSH concept, the MeSH category was also provided. The MetaMap summary was uploaded to R and a regular expression function was used to extract the noun phrases, mapping scores, candidate MeSH terms, and MeSH categories. MeSH terms were only included in the analysis if the mapping score was greater than or equal to 900. The frequency that each MeSH concept and category occurred in the HealthData.gov metadata was calculated.

Statistical Analysis

The R statistical software package was used to determine if the frequency of matching types (partial, exact, and no match) were significantly different between the mapped nouns and MeSH entry terms. The frequency of the partial matches was calculated as the number of nouns that matched at least part of a MeSH entry term. Exact matches were the frequency of nouns that exactly matched a MeSH entry term. The frequency of no matches was calculated as the sum of the nouns that did not appear in a MeSH entry term. Chi-squared tests of independence, with pairwise comparisons and a Bonferroni correction, were used to determine if there were significant differences in the frequency of terms in the matching groups.

Results

There were 30,132 nouns identified from 17,677 noun phrases from the 1,003 datasets using MetaMap, of which 2,975 nouns and 9,772 noun phrases were considered unique. When restricted to the MeSH library, there was a total of 18,839 concepts identified of which 1,336 were unique. The number of unique MeSH concepts that had a mapping score greater than or equal to 900 was 947. The 10 most frequent MeSH concepts and the accompanying categories that had a mapping score greater or equal to 900 are shown in Table 2.

Table 2 - Most frequent MeSH Concepts in HealthData.gov

Concepts	Categories	Frequency (Percentage)
Use (Utilization)	Functional Concept	238 (3.6%)
Medicare	Governmental or Regulatory Activity	134 (2.0%)
Collections	Intellectual Product	87 (1.3%)
Reported (Report)	Intellectual Product	84 (1.3%)
Programs	Intellectual Product	83 (1.3%)
File (Filing)	Occupational Activity	74 (1.1%)
United States	Geographic Area	74 (1.1%)
Medicaid	Governmental or Regulatory Activity	68 (1.0%)
Measures	Quantitative Concept	61 (0.9%)
Published (Publishing)	Occupational Activity	59 (0.9%)

Table 3 shows the frequency of nouns that matched partially, exactly, or did not at all match a MeSH entry terms. In addition, Table 3 indicates the frequency of these matching groups when the nouns were restricted to those covering 75% of all nouns, totally 480 nouns.

Table 3 - Matching HealthData.gov nouns to MeSH terms

Match Type*	Medical Subject Headings (MeSH)	
	All Terms Frequency**	*Restricted to top 75%* Frequency**
No Match	851 (28.6%)	47 (9.8%)
Partial	1533 (51.5%)	301 (62.7%)
Exact	591 (19.9%)	132 (27.5%)
Total	**2975 (100%)**	**480 (100%)**

**No match is the frequency where the noun is present in the HealthData.gov metadata, but not covered by the MeSH library. A partial match is the frequency of nouns that partially match a MeSH term. An exact match is the frequency of nouns that exactly match a MeSH term.*

*** There were significant differences in the frequency of each of the matching types (df=1; p-value<0.001).*

There were significant differences between the matching types (χ^2=477.3, df=2, p-value<0.001). Pairwise comparisons demonstrated that there were significantly fewer nouns that did not match to MeSH terms when compared to partial (χ^2=195.1, df=1, p-value<0.001) and exact matches (χ^2=46.9, df=1, p-value<0.001). There were significantly more nouns with partial matches to MeSH entry terms compared to nouns that exactly matched MeSH entry terms (χ^2=417.8, df=1, p-value<0.001).

These differences were observed when the nouns were restricted to the top 75% (χ^2=209.0, df=2, p-value<0.001). Pairwise comparisons demonstrated that there were significantly fewer nouns that did not match a MeSH terms compared to partial (χ^2=185.4, df=1, p-value<0.001) and exact matches (χ^2=40.4, df=1, p-value<0.001). Also, partial matches were significantly greater than exact matches (χ^2=66.0, df=1, p-value<0.001).

Discussion

There were significantly greater nouns that partially or exactly matched MeSH entry terms than nouns that didn't have any matches to MeSH entry terms. This is evidence that MeSH may offer adequate coverage of the concepts that are present in the HealthData.gov metadata. Although additional research is needed, this study suggested that MeSH may be a suitable controlled vocabulary to adopt for indexing HealthData.gov datasets.

As federal agencies and independent researchers acquire data, there is a general need to organize the data for later indexing and queries. One study found that very few publications that were registered at ClinicalTrials.gov actually published their datasets [31]. It was suggested that the lack of existing policies around standardization do not ensure availability of results and data from clinical research. Another study [32] discussed the need to develop data models and dictionaries that can support sharing of biomedical data, and suggested that the development of terminologies and ontologies could support this goal. Based on the results of the present experiment, the use of MeSH has the potential to support these demands for data sharing and data searching of a United States data portal, HealthData.gov.

Future research should evaluate if indexing with MeSH supports consumers in retrieving relevant datasets. If MeSH does offer a more effective search strategy, a method needs to be developed to automatically index the CKAN metadata. A limitation of this study is that new datasets are frequently published in HealthData.gov. This study only examined datasets that were published during a 2-year timeframe. Future research will need to be conducted to investigate the sustainability of coverage that MeSH. Finally, future research can investigate the term coverage of other terminologies, such as SNOMED CT.

Conclusion

This study demonstrated that MeSH may be a suitable vocabulary for indexing the data that is available through HealthData.gov. Better indexing would help the the Open Government Initiative, enacted by the United States federal government, achieve its goal of providing accessible published datasets. Adopting a standardized vocabulary such as MeSH for indexing the HealthData.gov metadata, the search and retrieval methods may be better supported. In order to ensure that MeSH can be used effectively for indexing, MeSH needs to have adequate coverage of the concepts in the HealthData.gov metadata. By extracting the nouns from the

metadata using MetaMap, a significant number of the nouns mapped partially or exactly to MeSH entry terms.

References

[1] Coglianese C. The transparency president? The Obama administration and open government. Governance 2009: 22(4): 152-67.

[2] Obama B. The Obama-Biden plan. Change.gov. 2007: [cited 2014 November 10]. Available from: http://change.gov/agenda/technology_agenda.

[3] Obama B. Transparency and open government. White House. 2009: [cited 2014 November 10]. Available from: http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment.

[4] Data.gov. 2014: [cited 2014 November 10]. Available from: http://www.data.gov.

[5] Orszag PR. Memorandum for the heads of executive departments and agencies. Executive Office of the President: Office of Management and Budget, 2009.

[6] OGP. The open government partnership: National action plan for the United States of America. White House. 2011: [cited 2014 November 10]. Available from: http://www.whitehouse.gov/sites/default/files/us_national_action_plan_final_2.pdf.

[7] Sunstein CR. Memorandum for the heads of executive department and agencies. Executive Office of the President: Office of Management and Budget, 2011.

[8] HealthData.gov. 2014: [cited 2014 November 10]. Available from: http://www.HealthData.gov.

[9] CKAN. About CKAN. 2014: [cited 2014 November 10]. Available from: http://ckan.org/developers/about-ckan.

[10] Napoli PM, and Karaganis J. On making public policy with publicly available data: The case of US communications policymaking. Gov Inform Q 2010: 27(4): 384-91.

[11] Evans AM, and Campos A. Open government initiatives: Challenges of citizen participation. J Policy Anal Manag 2013: 32(1): 172-85.

[12] Jaeger PT, and Munson S. Engaging the public in open government: Social media technology and policy for government transparency. Fed Reg, 2010.

[13] Dawes SS. Governance in the digital age: A research and action framework for an uncertain future. Gov Inform Q 2009: 26(2): 257-64.

[14] Dawes SS, and Helbig N. Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In: Wimmer MA, eds. Electronic Government: Lecture Notes in Computer Science: EGOV, 2010; pp. 50–60.

[15] Tapscott D, Williams AD, and Herman D. Government 2.0: Transforming government and governance for the twenty-first century. New Paradigm 2007: 1-21.

[16] Meijer A, and Thaens M. Alignment 2.0: Strategic use of new Internet technologies in government. Gov Inform Q 2010: 27(2): 113-21.

[17] Robinson D, Yu H, Zeller WP, and Felten EW. Government data and the invisible hand. Yale JL & Tech 2009: 11: 160-76.

[18] Chang A-M, and Kannan P. Leveraging Web 2.0 in government. IBM Center for the Business of Government, 2008.

[19] Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, et al. Open PHACTS: semantic interoperability for drug discovery. Drug Discovery Today 2012: 17(21-22): 1188-98.

[20] Scheufele E, Aronzon D, Coopersmith R, McDuffie MT, Kapoor M, Uhrich CA, et al. TranSMART: An open source knowledge management and high content data analytics platform. AMIA Joint Summits on Translational Science proceedings 2014: 96-101.

[21] Nelson SJ. Medical terminologies that work: The example of MeSH. Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms, and Networks: ISPAN, 2009; pp 380-4.

[22] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004: 32: 267-70.

[23] Aronson AR, and Lang FM. An overview of MetaMap: Historical perspective and recent advances. JAMIA 2010: 17(3): 229-36.

[24] Praz V, and Bucher P. CleanEx: New data extraction and merging tools based on MeSH term annotation. Nucleic Acids Res 2009: 37: 880-4.

[25] Uwimana E, and Ruiz ME. Integrating an automatic classification method into the medical image retrieval process. Annual Symposium proceedings: AMIA 2008; pp 747-51.

[26] Aronson AR, Mork JG, Lang F-M, Rogers WJ, and Neveol A. NLM Medical Text Indexer: A tool for automatic and assisted indexing. Bethesda: US National Library of Medicine, 2008.

[27] McCray AT, Srinivasan S, and Browne AC. Lexical methods for managing variation in biomedical terminologies. Proceedings of the Annual Symposium on Computer Application in Medical Care: AMIA 1994; pp 235-239.

[28] Aronson AR. The effect of textual variation on concept based information retrieval. Annual Symposium proceedings: AMIA 1996; pp 373-377.

[29] Aronson AR. MetaMap Evaluation. Bethesda: US National Library of Medicine, 2001.

[30] NLM. Medical Subject Headings. NLM. 2014 [cited 2014 November 10]. Available from: https://http://www.nlm.nih.gov/cgi/request.meshdata.

[31] Shamliyan TA, and Kane RL. Availability of results from clinical research: Failing policy efforts. JEGH 2014: 4(1): 1-12.

[32] Thibault JC, Roe DR, Facelli JC, and Cheatham III TE. Data model, dictionaries, and desiderata for biomolecular simulation data indexing and sharing. J Cheminformatics 2014: 6(1): 1-23.

Address for correspondence

David T. Marc – marcx003@umn.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-534

Recruit - An Ontology Based Information Retrieval System for Clinical Trials Recruitment

Diogo F. C. Patrão[a], Michel Oleynik[a], Felipe Massicano[a], Ariane Morassi Sasso[a]

[a] *International Center for Research (CIPE) - A. C. Camargo Cancer Center*

Abstract

Clinical trials are studies designed to assess whether a new intervention is better than the current alternatives. However, most of them fail to recruit participants on schedule. It is hard to use Electronic Health Record (EHR) data to find eligible patients, therefore studies rely on manual assessment, which is time consuming, inefficient and requires specialized training. In this work we describe the design and development of an information retrieval system with the objective of finding eligible patients for cancer trials. The Recruit system has been in use at A. C. Camargo Cancer Center since August/2014 and contains data from more than 500,000 patients and 9 databases. It uses ontologies to integrate data from several sources and represent medical knowledge, which helps enhance results. One can search both in structured data and inside free text reports. The preliminary quality assessments shows excellent recall rates. Recruit proved to be an useful tool for researchers and its modular design could be applied to other clinical conditions and hospitals.

Keywords:

Information Retrieval, Ontology, Clinical Trial, Patient Selection

Introduction

A clinical trial is designed every time a new intervention needs to be tested and compared with its current alternatives in an effective way. This type of study usually specifies two groups of research subjects: control (patients receiving regular treatment or placebo) and treatment (patients receiving new treatment). To statistically back up the conclusions, there should be enough participants (the recruiting target) and all of them should have similar clinical characteristics (in order to meet the inclusion and exclusion criteria). Currently, only half of all clinical trials reach their recruiting target, and half of those reach the target in a timely manner [1]. The usage of existing search tools (such as data warehouses) over Electronic Health Record (EHR) data can lead to significant increases in recruitment rates [2–5], however, there are still important challenges to be faced.

Firstly, the clinical data is notoriously difficult to represent formally [6–8]. The complex nature of medical practice, its specialties and sub-specialties led the industry to develop specialized software. Thus, it is common that the hospital context involves usage of several systems for different aspects of clinical practice, and the patient information is scattered through different databases, usually using different data models. Both structured forms and reports containing free text may coexist in one database. Therefore, in order to use EHR data to find patients based on clinical features, it is necessary to apply data integration techniques [9].

In Computer Science, ontologies represent knowledge formally, and have been used to derive new knowledge from facts by a process called logical inference [10]. Ontology based data integration employs the descriptive power of ontologies to harmonize semantic mismatches between source databases, particularly on the biomedical domain [11]. Additionally, the field of Information Retrieval has been developed at a fast pace [12], with continuous improvements to the algebraic models (such as latent semantic indexing) as well as to the probabilistic models, many of them empowering natural language processing. However, the availability of language models trained on the medical subdomain is still sparse mainly due to ethical concerns when publishing corpora, a fact that hinders the development of the area.

Several systems have tried to overcome these difficulties with customized search engines. Most of them [13] map free-text data onto concepts taken from terminologies such as SNOMED CT, upon which a structured search is then performed. The translation of eligibility criteria into computable definitions is usually done by the searcher, but can be aided through machine language [14,15], especially useful on cross-language environments.

In this paper, we present Recruit, an information retrieval system with the target of finding patients based on clinical criteria using EHR data. It uses ontologies to reconcile heterogeneous databases, as well as merge data from structured forms and free text for pathology and image reports. We will describe its requirements, design, deployment and preliminary quality assessments.

Materials and Methods

Requirements

We classified requirements for Recruit in two types: data and functional. Data requirements are clinical information that the system must be able to find, and functional are tasks that the user must be able to perform when using the system. The first source for data requirements was the Medical Informatics Laboratory ticket system. We analyzed the patient reports requested during the past 12 months and identified the most common clinical criteria. We interviewed the clinical trial support team, who described the most important concepts.

In particular, the system should implement the ethics regulations regarding patient data access. System usage should be monitored by analytics tools, to assess usage patterns.

Design

Based on the verified requirements, we elaborated an UML use case diagram to describe the interactions between the user

and the system, and a mock-up of the user interface. File formats and software used should be open source.

We chose to use ontologies to represent both patient information and knowledge about the domain. The data model should be as simple as needed for creating an index of concepts for patients. Existing ontologies can be reused, as long as they are provided with a free license and are within the desired expressive power. Classes and properties will be annotated in Portuguese, to allow easy understanding to both the technical team and Recruit users.

Quality Assessment

In order to evaluate the quality of Recruit we created a methodology to build gold standards that could be replicated and that would help us measure the impact of interventions in the system. The standards are constructed based on the classification of the results from predetermined queries as relevant or non-relevant in a chosen scenario. Therefore, we can obtain the precision (defined as the ratio of retrieved patients that are relevant) and recall (defined as the ratio of the relevant patients that were retrieved) measures for each one, which combined generates the weighted harmonic means F_1 and F_2. These scores allow us to assess the quality of the result sets produced by Recruit.

First, we realized that it is important to choose queries in which the professionals from the institution have scientific interest for and also that brings a reasonable amount of cases. In this way, it is possible to collaborate with researchers in the construction of the gold standards and the results can be assessed in a short period of time.

Second, each result from the chosen queries had to be classified as relevant or non-relevant, by using a 3 point scale in which (2) means relevant, (1) unsure and (0) non-relevant. Therefore, if there is certainty about the presence of the clinical condition, we would classify the subject as relevant (2), usually by confirmation from a pathology report. Besides, if there is doubt, either because there is lack of electronic clinical documents to support the presence of the condition or, by the own doubt of the health professionals, we would classify the case as unsure (1). Finally, if the hypothesis of the patient meeting the clinical criteria is rejected by the physicians or, if it is impossible to find the information even in non-electronic format (e.g. the case is too old), we would classify the case as (0) non- relevant.

The next step is to convert the 3 point scale into a binary one, thus we prefer to consider only relevant cases as (1) and the all others as non relevant (0). Finally, it is recommended to look into documents not searched by the system to recover and analyze other cases related to the queries. That will allow the generation of the described set of measures for each standard, namely precision (P), recall (R), F_1 and F_2, which will further help the system's quality evaluation.

Results

Requirements

After the interview with the clinical trials support team and review of requests on the ticket system, we compiled the data requirements (Table 1). After that, we identified in which database each information could be found and Table 2 compiles the characteristics of each one.

Table 1 – Data Requirements

% Tickets	Description
33,07%	Diagnosis (ICD-10) , date and age of diagnostic
33,07%	Tumor Topography (ICD-O)
33,07%	Tumor Morphology (ICD-O)
13,39%	Treatment type (Surgery, Chemotherapy, Radio-therapy)
13,39%	Sample collected in biobank (by type: frozen tissue, blood, RNA, DNA)
11,02%	Chemotherapic drugs used
9,45%	Demographics (Date of birth, current age, gen-der)
8,66%	Department
6,30%	Cancer status (under diagnostic, not cancer, cancer, metastasis, remission)
6,30%	Alive/deceased, date of last information, death by cancer or other reasons
4,72%	TNM Clinical Stage (e. g. CS III)
3,94%	TNM Variables (e. g. T0, N2, M1)

Table 2 – Source Databases

Database	Technology	Information	Time span	Structured / free text
EHR	Oracle 10i	TNM, ICD-10, Chemotherapy, Surgery, Radiotherapy	2007-2014	Structured
AP Reports	Interbase	Pathology Reports	2006-2014	Free text
AP Reports-legacy	Oracle 8i	Pathology Reports	2000-2006	Free text
AP Reports-legacy-2	DBase III	Pathology Reports	1992-2000	Free text
Image Reports	Oracle 10i	Image Reports	2006-2014	Free text
Image Reports-legacy	Oracle 8i	Image Reports	2002-2006	Free text
Registry	Excel	ICD-10, ICD-O, TNM, follow up, treatment	2000-2012	Structured
Biobank	MySQL 5.1	Frozen tissue, blood, DNA/RNA sample	1999-2014	Structured
Index	SPSS	ICD-10, ICD-O	1953-2000	Structured

Design

The Recruit software is divided in two parts: the backend, responsible for extracting, transforming and indexing data; and the frontend, the web interface with which the end user interacts. We designed a single use case (depicted on Figure 1) to represent the user authentication, search and retrieval of results on screen or on a CSV file.

The backend is a workflow that produces an index combining unstructured and structured data contained in the original databases. This index will be processed by the search engine Apache Solr which will be queried by the frontend. We adapted the structure described in [16]: instead of view integration, we created a data warehouse based on ontology, and then loaded this data as structured metadata in an indexing server, along with unstructured report texts.

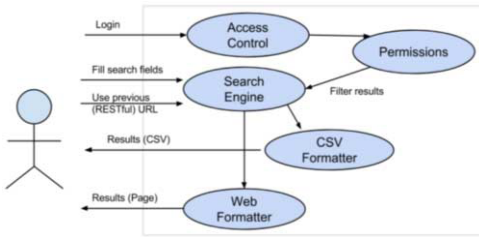


Figure 1 - UML use case diagram

As the first step, we integrated data from several structured databases into only one triple store endpoint, namely Openlink Virtuoso (Figure 2). The mapping from relational data to RDF graphs was done with the help of Ontop, thus generating several RDF files which were loaded in memory using ARQ (configured with inference profile OWL Micro Reasoner). Then, all inferenced triples were materialized into a single RDF file, later loaded into the triplestore endpoint. Finally, a set of handcrafted SPARQL queries was executed at the endpoint in order to enrich it with information not feasible by the inference process due to absence of closed world reasoning. Therefore, the resulting dataset contained consolidated information, such as the ICD-10, ICD-O and the date of the diagnosis and also the age of the patient at the diagnosis.

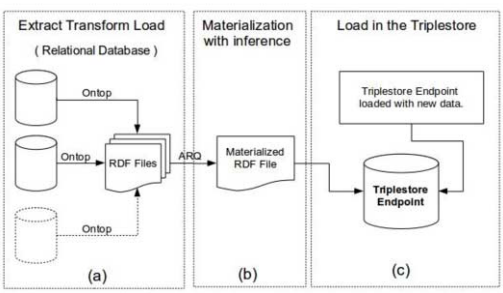


Figure 2 - Integration process data flow

In a second step, we published the inference-rich facts together with unstructured data into a search engine (Figure 3). We extracted the textual content of pathology and image reports and then associated it with the triple store endpoint data using the patient identifier. The results of this process were published in the search engine using its API.

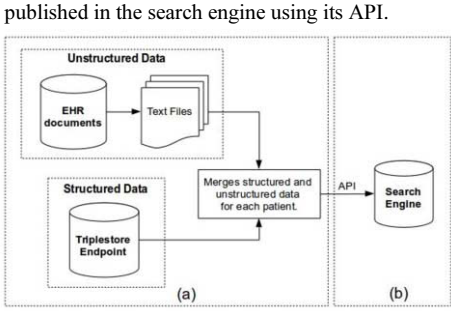


Figure 3 - Search engine publishing data flow

Knowledge representation and structured data mapping

We decided to represent structured data as an ontology, as it has been used for database integration and semantic harmonization, and also for its inference capabilities. This consolidated data will be used to feed the search engine and to display result details in the user interface. Due to the large data volume, we decided to use a very simple model, that would allow for a reasonably fast (for weekly updates) inference, while powerful enough to make required reasoning tasks.

In our ontology, patients are represented as instances of NamedIndividuals (as defined in OWL2 [17]) and structured search criteria as classes directly attributed to instances. For instance, a patient with breast cancer would belong to class C50 (the ICD-10 [18] code for this disease). The patient URI is based on the patient identifier, used in all databases. The ICD-10 ontology was reused from NCBO [19], the ICD-O was semi-automatically obtained from an existing relational database. Other ontologies (treatments, chemotherapy, cancer staging [20]) were manually created. Data properties were created for identification data and diagnostic dates. Properties with unique values for a patient were marked as *owl:functional* to allow sorting on them. All properties and classes were created with URIs in Portuguese and annotated with the *skos:prefLabel* property. Besides the class hierarchy, inference axioms were created to enhance results, specially on clinical staging data. See excerpts of the ontology axioms on Figures 4 and 5.

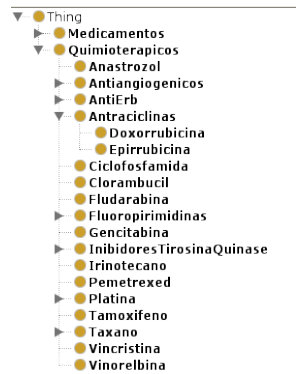


Figure 4 - Class hierarchy for chemotherapic drugs

SQL Mapping

In order to translate data from the relational databases into an ontology, we created one SQL query that maps a production database to instances of the corresponding class. We used Ontop software to test and extract data into N3 format. Similar information represented in different ways in the same database was normalized in this step using regular expressions (such as '^IV|IV$|^4$' for clinical stage IV). Moreover, we associated a patient ICD-10 code with his/her diagnostic status, a structured field with three possible values: patient in diagnostic, patient without cancer, or patient with a confirmed cancer diagnostic.

Figure 5 -Rules for inference of TNM clinical stage for specific topographies

Quality Assessment

The first query we chose to apply our methodology on was related to the phyllodes tumor, a rare breast neoplasm. With the help of a research nurse we built a list of synonyms in Portuguese for that clinical term and they were expressed in the following query: "filode OR filoides OR phyllodes OR phylloides OR phylodes OR phyloides". Then, we presented it to Recruit, which retrieved a set with 266 patients and, to it we added more 91 that were found by looking into documents not searched by Recruit but that belong to the patients' electronic health record. That was done in order to reproduce a more real search universe and finally, we divided the cases between two authors of this paper and one research nurse, who classified them as relevant (2), non-relevant (0) or unsure (1). There was a final meeting to guarantee that everyone had the same understanding about each classification option and so, the measures highlighted in table 3 were obtained. The evaluated recall is high (97.16%), which fulfill the health professionals expectations. However, the obtained precision was 51.50%, probably due to the search engine not being specific for the healthcare scenario. F_2 (82.53%) is higher than F_1 (67.32%), since the former gives more weight to recall while the latter considers both measures to have the same weight.

Table 3 – Quality Evaluation

Variable	(%)
Precision	51.50
Recall	97.16
F_1	67.32
F_2	82.53

We noted some signs that may indicate whether a case is relevant or not, for example, if the patient has an unknown name or if the case is too old, it usually is less relevant due to lack of electronic records and the possibility of the patient being dead or have lost follow-up. It is also a sign of non-relevance the presence of terms that negates the searched condition such as 'absent' or the existence of relevant documents only in paper or microfilm. If the patient possesses a considerable amount of electronic records, or if the clinical criteria appears in an anatomic pathology report and the latter is recent, the chance of the case being relevant is higher.

Discussion

We designed Recruit to be easily maintainable, modularizing the data extraction in a large number of queries, each mapping to specific classes, or group of classes within the same category or property, on a particular database. Adding another database or removing a problematic query simply implies adding or removing a mapping file. We verified this in practice, when developing the mapping files incrementally.

Our database is large (more than 500,000 patients), and our mappings generated a reasonable amount of RDF triples which took a significant time (>6 hours) to be extracted. Most of this time was spent on a couple of queries that used regular expressions to detect patients with specific molecular test results. The inference materialization step took more than 2 hours and required almost 20GB of RAM in a single process. This could be optimized by using a customized program to extract specific axioms, such as class instances, instead of using a general SPARQL query to extract all possible inferences. Parallel inference is a current topic of study, and could greatly enhance the inference performance. Overall, the preparation step took more than 8 hours, hindering the application to be updated on a daily basis. The addition of concepts, axioms or mappings should be carefully planned, as running time and memory required can grow exponentially.

Our modelling also directly connects the diagnostic to the patient. Although the inference takes less time this way, it would be impossible to correctly assign the date of a diagnostic when a patient has multiple tumours. The creation of a new entity for the diagnostic would allow this, but the inference resources would be higher, and it would require a different usage of the indexing engine as well.

Since its release on August/2014 until December/2014, the system has been continuously accessed by at least 15 different users per week. Therefore, it is becoming a established tool for researchers in A. C. Camargo Cancer Center.

Recruit achieved a high recall rate, considered good by the research nurses, since they usually look into all the results, thus not giving much weight to precision. Although we understand the importance of the former measure, the latter has a significant role, since it can diminish the time a researcher needs to look for relevant patients. Also, the search engine being used by Recruit is capable of dealing with cases of inflection in Portuguese but not specificities from the healthcare field. We believe that Recruit would have its performance enhanced if it was imbued with a clinical health terminology in Portuguese, which would allow users to discover the synonyms for terms such as "phyllodes".

Despite having just one gold standard, we believe that it gave us some insight related to the quality of the search and to the signs needed to enhance it. Therefore, if we build more standards to evaluate the searches made by Recruit we can improve our list of signs and use it to make interventions in the system, which will have its results measured and compared to the standards. Also, each intervention has a different difficulty level: in order to increase the weight given to terms found in anatomic pathology reports, for example, one would need to detect if the searched condition is followed by a word or phrase that negates it. All that must be taken into consideration in order to plan the next maintenance projects and consequently improve Recruit's quality.

We could have used extraction and classification techniques to improve free-text data with structured concepts and thus improve recall rates [21]. However, due to unavailability of public language models and annotated corpora in the medical domain in Portuguese, we chose not to apply experimental results at this moment while further research is ongoing.

Other works [3–5] rely on search and integration tools already implemented for performing the search; in particular, these tools were designed with other objectives than searching for patients for clinical trials, therefore they may not include all needed search criteria, or may not be customized to do so and keep their main objective. Recruit, being designed exclusively for patient selection, can be customized as needed for the sole purpose of finding potentially recruitable patients. Also, as it relies on open source software and file formats, it does not impose artificial limits or require expensive licenses, which is particularly relevant for research projects. Recruit integrates data from structured and unstructured sources, which, at least to our knowledge, has not been done in this field.

Recruit is now a production level tool being used at A. C. Camargo Cancer Center, but further improvements are

needed. A comprehensive set of gold standards should be created in order to guide further enhancements on mappings, and evaluate the general quality of results. Data extraction and inference should be optimized as to be more frequently updated. The ontology model should be improved by creating an instance for each diagnosis and relating the diagnostic classes and date to it. This would allow specifying date of diagnostic for patients with multiple cancers. Also, other types of documents should have free text indexed, such as outpatient clinical notes. A preliminary evaluation showed that doctors employ a large number of acronyms and do not use proper phrases, instead using a list of diagnostics, therapies and other applicable concepts, and that should pose a major challenge.

Conclusion

We have successfully implemented an ontology based information retrieval system for clinical criteria based patient selection. It uses EHR data, represents medical knowledge as ontologies, integrates several databases and allows search for structured data and free text. The preliminary quality assessments show excellent recall rates. It is not only an important asset for A. C. Camargo Cancer Center researchers, but the principles here presented can be used on a larger range of information retrieval problems.

Acknowledgments

Marcelo Sagayama contributed on the prototype construction, data mapping and other technical tasks. Fábio Rampazzo Mathias provided technical assistance. Helano Carioca Freitas and Tatiana Iafuso provided system requirements. Silvana Soares dos Santos assisted on the gold standard construction.

References

[1] Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S. Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. BMJ Open. 2012; 2(1):e000496.

[2] Fink E, Kokku PK, Nikiforou S, Hall LO, Goldgof DB, Krischer JP. Selection of patients for clinical trials: an interactive web-based system. Artif Intell Med. 2004; 31(3):241–54.

[3] Dugas M, Lange M, Müller-Tidow C, Kirchhof P, Prokosch H-U. Routine data from hospital information systems can support patient recruitment for clinical studies. Clin Trials. 2010; 7(2):183–9.

[4] Kamal J, Pasuparthi K, Rogers P, Buskirk J, Mekhjian H. Using an information warehouse to screen patients for clinical trials: a prototype. AMIA Annu Symp Proc. 2005; 5(6):1004.

[5] Köpcke F, Kraus S, Scholler A, Nau C, Schüttler J, Prokosch H-U, et al. Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition. Int J Med Inform. 2013; 82(3):185–92.

[6] Beale T. The Health Record - why is it so hard? In: Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2005. Stuttgart: Schattauer; 2005. p. 301–4.

[7] Jaspers MWM, Knaup P, Schmidt D. The Computerized Patient Record: Where Do We Stand? 2006; 29–39.

[8] Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: A literature review. Journal of Biomedical Informatics. Elsevier; 2010. p. 451–67.

[9] Anjum A, Bloodsworth P, Branson A, Hauer T, McClatchey R, Munir K, et al. The Requirements for Ontologies in Medical Data Integration: A Case Study. 11th International Database Engineering and Applications Symposium (IDEAS 2007). IEEE; 2007. p. 308–14.

[10] Brachman RJ, Levesque HJ. Knowledge representation and reasoning. New York. Morgan Kaufmann; 2004.

[11] Sujansky W. Heterogeneous database integration in biomedicine. J Biomed Inform. 2001; 34(4):285–98.

[12] Manning CD, Raghavan P, Schütze H. An Introduction to Information Retrieval. Press, Cambridge U.; 2008.

[13] SEMCARE [Internet]. 2015 [accessed 2015 Apr 10]. Available from: http://semcare.eu/

[14] Markó K, Schulz S, Hahn U. MorphoSaurus - design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. Methods Inf Med. 2005; 44(4):537–45.

[15] Luz FF. Querying ontologies using controlled natural language. Universidade de São Paulo; 2013.

[16] Patrão DFDC. Desenvolvimento e Avaliação de Ferramentas Computacionais para Triagem Automática de Sujeitos de Pesquisa. A. C. Camargo Cancer Center; 2014.

[17] W3C. OWL 2 Web Ontology Language Primer [Internet]. 2009 [accessed 2014 Aug 21]. Available from: http://www.w3.org/TR/owl2-primer/

[18] World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD-10). World Health Organization; 2004.

[19] Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story M-A, et al. The National Center for Biomedical Ontology. J Am Med Inform Assoc. 2012; 19(2):190–5.

[20] Sobin L, Gospodarowicz M, Wittekind C, editors. TNM Classification of Malignant Tumours. John Wiley & Sons; 2011.

[21] Oleynik M. Clinical reports information retrieval. Universidade de São Paulo; 2014.

Address for correspondence

Diogo Ferreira da Costa Patrao

International Center for Research (CIPE)

A. C. Camargo Cancer Center

djogo@cipe.accamargo.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-539

Assessing the Need of Discourse-Level Analysis in Identifying Evidence of Drug-Disease Relations in Scientific Literature

Majid Rastegar-Mojarad [a,b], Ravikumar Komandur Elayavilli [a], Dingcheng Li [a], Hongfang Liu [a]

[a] *Biomedical Statistics & Informatics, Mayo Clinic, Rochester, MN, USA*
[b] *University of Wisconsin-Milwaukee, Milwaukee, WI, USA*

Abstract

Relation extraction typically involves the extraction of relations between two or more entities occurring within a single or multiple sentences. In this study, we investigated the significance of extracting information from multiple sentences specifically in the context of drug-disease relation discovery. We used multiple resources such as Semantic Medline, a literature based resource, and Medline search (for filtering spurious results) and inferred 8,772 potential drug-disease pairs. Our analysis revealed that 6,450 (73.5%) of the 8,772 potential drug-disease relations did not occur in a single sentence. Moreover, only 537 of the drug-disease pairs matched the curated gold standard in Comparative Toxicogenomics Database (CTD), a trusted resource for drug-disease relations. Among the 537, nearly 75% (407) of the drug-disease pairs occur in multiple sentences. Our analysis revealed that the drug-disease pairs inferred from Semantic Medline or retrieved from CTD could be extracted from multiple sentences in the literature. This highlights the significance of the need of discourse-level analysis in extracting the relations from biomedical literature.

Keywords:
Relation extraction; Discourse-level analysis; Literature-based discovery; Semantic Medline.

Introduction

Information extraction (IE) aims to automatically extract information from text. To understand and extract information from text, IE systems should look at the text as whole and not only sentences separately. Researchers in the discourse domain agree that usually sentences/clauses are not understood in isolation [1]. Researchers in IE have studied discourse-level analysis [2] in applications such as question answering and dialogue generation. Of course, discourse-level analysis is built on top of sentence-level analysis and its performance somehow depends on sentence level. One of the well-studied parts of discourse-level analysis is identifying relation between sentences (clauses), called discourse relation. These types of relations could be contrast or explanation-evidence [1].

Two main tasks in IE are named entity recognition (NER) and relation extraction. In relation extraction, IE systems identify the relation between two or more entities where the entities could be located in one or multiple sentences. Identifying the latter type of relations needs discourse-level analysis such as coreference resolution which can be challenging.

The goal of this paper is to asess the need of discourse-level analysis for relation extraction and indirectly evaluate the amount of information that IE systems are missing if they just focus on sentence level. First, to perform this experiment, Semantic Medline [3] is used to generate a list of potential literature-based discoveries. The potential discoveries are limited to drug-disease relations as potential drug repositioning candidates. Then, we implement two kinds of searches on top of Medline abstracts to find any evidence of these discoveries, 1) Sentence-level search: we looked for pairs in the sentence 2) Discourse-level search: the whole abstract was searched. Comparing the results of these searches indirectly evaluates the amount of relations that IE systems can extract from sentence-level v.s. discourse-level analysis.

Background

Relation Extraction

The IE pipeline usually begins with NER which has been very well studied and various types of methods such as dictionary-based, rule-based, and machine learning algorithms have been applied to NER [4], [5]. After NER, the main task in IE is identifying relations among the entities. Relations could be binary or they could involve more than two entities. Most of relation extraction systems focuson relations in a single sentence [6]. It is imperative for a relation extraction system to extend beyond clauses and sentences and handle complex discourse-level analysis.

The importance of discourse-level analysis is well known and the analysis has been studied in several applications such as: question-answering systems [7]–[9], automatic dialogue generation [10], etc. Besides, there are several studies on extracting discourse-level relations. Marcu and Echihabi [1] developed an unsupervised approach to classify four types of relations between sentences/clauses. The goal here was extracting the relation between two or more entities that are mentioned in multiple sentences. Bach and Badaskar [6] reviewed some of these types of relation extraction systems and their applications. The methods for extracting such relations, often range from a simple distance based criteria to a more complex one that employs statistic and linear algebraic approaches [11] to extract explicit and implicit semantic relations from text.

In the biomedical domain, researchers developed various IE systems to extract different types of biomedical relations [12]–[16]. Quan et al [13] proposed two systems, unsupervised and semi-supervised to extract protein-protein interactions and

gene-suicide associations. Their systems employed dependency parsing. Bundschus et al [14] developed biomedical relation extraction using conditional random fields. Their system identified relations between diseases and treatments; also relations between genes and diseases. All these systems just focused on relations within a single sentence and often ignored relations involving entities across sentences. In general, extracting relations that involve entities mentioned in two separate sentences is a complex one and requires special NLP techniques such as coreference resolution and complex semantic analysis.

Literature-based discovery

Literature-based discovery (LBD) aims to find a connection/correlation between concepts using scientific literature. Therefore, LBD is one kind of special relation detection we are interested in. Many LBD systems have been developed in the biomedical domain to generate new hypotheses that potentially could lead to new discoveries. Swanson [17] first introduced this approach and applied it to find a correlation between migraine and magnesium. After that, a number of studies followed his approach and had interesting discoveries. The important part of LBD is how to decide that two concepts are correlated. The most commonly used approaches are co-occurrence analysis [18], Association Rules [19], TF-IDF, Z-Score, and Mutual Information Measure [20]. Yetisgen-Yildiz and Pratt [20] briefly discussed these approaches and meanwhile Andronis et al. [21] reviewed literature mining systems which identify potential drug repurposing candidates. Another approach to identify correlated concepts is using semantic relations [22]. Hristovski et al. proposed [22] using semantic predications to enhance LBDs. In this study, semantic predications are used to create a list of potential LBDs (drug-disease pairs). The generated LBDs in this study are potential drug repurposing candidates.

Semantic Medline, semantically enhanced with predicates extracted by SemRep [23] from Medline titles and abstracts, contains approximately 70 million semantic predications. Predications are triplets of *Subject*, *Predicate*, and *Object* where the subject and object are biomedical entities (drug, gene or disease in our study), and predicate shows the type of relation between the entities such as inhibits, interaction with, associated with, etc. The predications are stored in a relational database called Semantic Medline Database (SemMedDB). It has been used in many studies to facilitate knowledge discovery [24], [25].

In this study, we plan to assess the need of going beyond sentences for extracting relations. Our goal in this study is to highlight the need and asess the amount of relations, relation extraction systems that are potentially missing if they just focus on sentence-level analysis. For this purpose, following Swanson's model, a list of potential drug-disease relations is generated from literature. The extracted LBDs are assumed as relations between drugs and diseases that could be mentioned implicitly or explicitly in literature. Two kinds of searches are conducted to identify any evidence of these relations in the sentence level or discourse level of Medline abstracts.

Methods

Our study contains two steps: 1) generating a list of drug-disease pairs based on LBDs and 2) using the discoveries to evaluate the drug-disease relation extraction by comparing the extractions from a single sentence to the one from multiple sentences.

Generating LBDs

In the first step, we generated a list of potential relations between drugs and diseases. We followed Swanson's model [17] to generate a list of LBDs. According to Swanson's model if one scientific study notes a correlation between concept A (Starting concept) and concept B (Linking concept), and another study mentions a correlation between concept B and concept C (Target concept), then there might be a correlation between concept A and C. In our study, drug is the starting concept and disease the target concept, gene serves as a conceptual link between the two leading to the generation of a list of potential drug-disease relations.

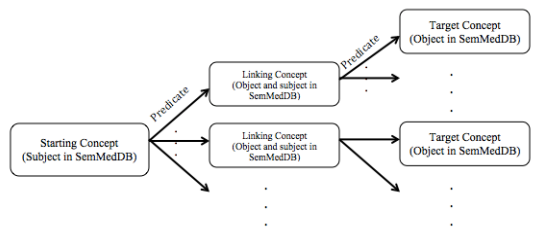


Figure 1 - Architecture of our LBD system. In this system, the starting concept is drug, the linking concept is gene, and the target is disease that leads to drug-disease discoveries. Our system uses Semantic predications as evidence of correlation between the concepts.

In contrast to commonly used approaches such as document-level or sentence-level co-occurrence of entities or concepts to generate LBDs, we used semantic predications as evidence for correlation between entities [22]. For example, from the following two predications from SemMedDB:

- Flecainide (Drug) *INTERACTS WITH* SCN5A (Gene)
- SCN5A (Gene) *ASSOCIATED WITH* Heart Failure (Disease)

The system generates *Flecainide-Heart Failure* as a potential new drug-disease pair. Figure 1 shows the architecture of our LBD process.

There are two major reasons for our choice of using semantic predication to generate LBDs 1) semantic predication takes biological meaning into consideration 2) the semantic type of the interaction and the contextual information about the interaction such as *NEGATION* allow us to filter unnecessary predications: *NEGATIVE TREATS*, *NEGATIVE ASSOCIATED WITH*, etc. At this point, we do not consider any threshold based on a frequency measure to further eliminate drug-disease pairs. We believe that using a frequency measure will eliminate potentially novel relations from our initial list.

In order to further refine the drug-disease pairs to be relevant for LBD, we used Timeline profiles to narrow down the drug-disease pairs with literature evidence. We only considered drug-disease pairs where the first literature evidence appeared after their related drug-gene and gene-disease pairs. Equation (1) clarified our Timeline profile:

$$Max\begin{pmatrix} Min(Y_{sp}(Drug-Gene)), \\ Min(Y_{sp}(Gene-Disease)) \end{pmatrix} \leq Min(Y_c(Drug-Disease)) \quad (1)$$

where Y_{sp}(Drug-Gene) indicates the publications' date of all Medline abstracts that contain at least one semantic predication between Drug-Gene. For Drug-Disease, we considered publications that contain at least one co-occurrence of the entities. For example, assume two studies in 2003 and 2005 reported an association between Drug A and Gene B; and three studies in 2006, 2008, and 2009 mentioned an association between Gene B and Disease C. The left hand side of the equation would be Max (Min (2003, 2005) , Min(2006, 2008, 2009)) = 2006. So if Drug A and Disease C appeared together in any publication before 2006, we do not consider Drug A-Disease C as a potential discovery.

Evaluating sentence and discourse-level relation extraction

After generating LBDs, we used the drug-disease pairs identified earlier to evaluate their co-occurrence in literature, thereby estimating the need for discourse-level analysis in relation extraction. So, we have a list of drug-disease relations that potentially could be mentioned in the scientific literature. In the second step, we searched Medline abstracts tofind any co-occurrence (evidence) of drug and disease pairs. We categorized the drug-disease pairs with at least one evidence in Medline, into two groups, i) sentence-level relation and ii) discourse-level relation. For a given pair, if the drug and the disease appeared in a single sentence in at least one of the abstracts, we classified it as a sentence-level relation, otherwise we considered it as a discourse-level relation.

In order to assess the true validity of the drug-disease pairs we used the Comparative Toxicogenomics Database (CTD) [26], manually curated biological relations as our reference standard. CTD contains annotations of biological relations from various categories such as chemical-disease, gene-disease, drug-gene associations along with the corresponding PubMed citation. We compared the drug-disease pairs identified by the system against the chemical-disease associations in CTD. We considered our pair to be valid only if there was a match in the PubMed citation in addition to the drug-disease pairs.

Results

Retrieval of LBD relations

In order to generate potential LBD pairs, we started with 1,710 approved drugs from DrugBank [27]. We extracted 4,096 drug-gene (A-B) unique pairs where the drugs were restricted to our chosen list of 1,710 along with the semantic predicates from SemMedDB. For all the genes mentioned in A-B pairs we further retrieved 2,741 gene-disease (B-C) relations from SemMedDB. With gene being the common link between Drug-Gene and Gene-Disease, we inferred 71,842 drug-disease (A-C) relations. Further analysis revealed that only a small fraction of (118) of the 71,842 drug-disease pairs had an overlap in terms of abstracts from which the drug-gene (A-B) and gene disease (B-C) pairs were identified. We also found 14,451 drug-disease pairs catalogued in SemMedDB since they co-occurred in the same sentence. We have 57,391 drug-disease pairs also indirectly inferred through the Swanson's model of LBD and not present in SemMedDB. The results of our study is illustrated in Figure 2.

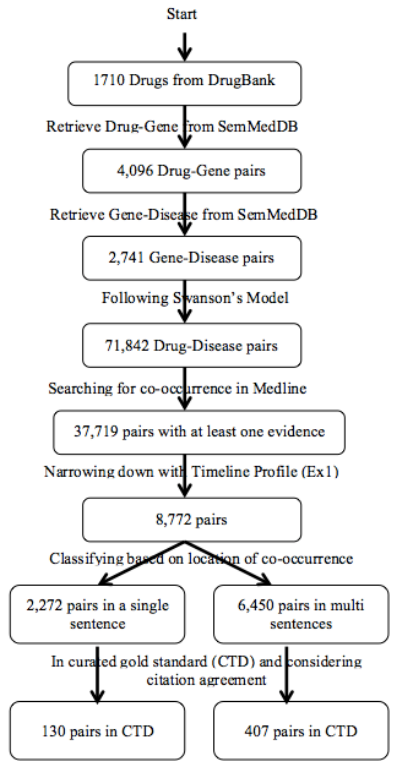


Figure 2 – Results of our study

Comparison of sentence level and discourse level

We further attempted to assess the level of textual extraction required to identify the drug-disease relations (A-C pairs). We searched Medline using "Drug AND Disease" as the query, to find any evidence of the identified pairs in the first step. We found only 37,719 (52.05%) of the 71,842 drug-disease pairs with at least one literature evidence.

Timeline analysis (Eq1) further narrowed down the number of drug-disease pairs to 8,772 (23.25%) from 37,719. 6,450 drug-disease relation pairs (73.52%) out of 8,772 identified earlier transcend sentence boundaries, demanding the requirement of discourse-level analysis for textual extractions.

We performed additional analysis on the remaining 2,322 drug-disease relational pairs, which had at least one literature evidence as sentence-level co-occurrence. For these 2,322 pairs, we found 89,805 literature evidences, which indicated that there were more than one literature evidence for each pair. Further composite analysis revealed that there were far more literature evidences across sentences than from a single sentence as shown in Figure 3.

We further carried out an assessment to validate the 8, 322 drug-disease pairs against a curated resource (CTD). We found that only 537 (6.4%) of them matched the gold standard. The low match was due to the fact that in addition to the match in the drug and disease names we also considered an agreement in the cited literature. Further analysis revealed that only 130 (24.20%) of the 537 that matched the gold standard occurred in a single sentence while the rest (75.80%) appeared across different sentences. If we ignored the match in the literature citation then we found 17,094 drug-disease pairs in CTD, which was significantly higher than the earlier one.

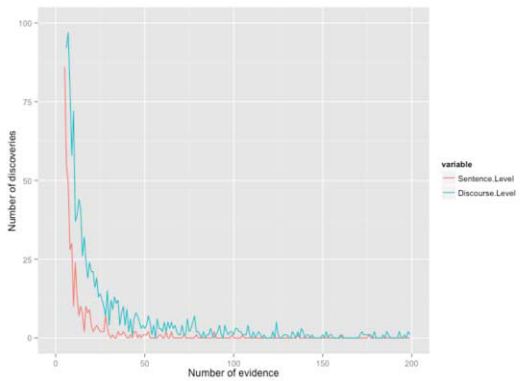


Figure 3 – Comparison of frequencies of drug-disease relations' co-occurrence in a single sentence versus multiple sentences

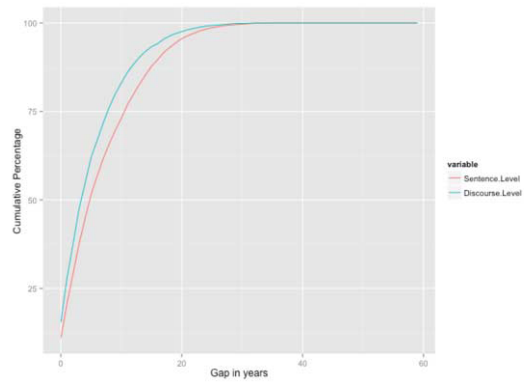


Figure 4 - Comparing the time gap between the first co-occurrence of discovery and the causal pairs (Sentence level versus discourse level)

Discourse-level analysis may impact Time lag of LBD

We performed another interesting analysis to study whether discourse-level analysis would have a positive impact on the time lag due to the reporting of causal pairs (drug-gene/gene-disease) and the appearance of drug-disease pairs in the scientific literature. Figure 4 plots the cumulative percentage which represents a cluster of drug-disease pairs observed at a time zone at both sentence level and discourse level. The trend shows that performing discourse-level analysis would significantly advance the identification of the drug candidate for a specific disease.

Discussion

In this paper, we attempted to assess the need for performing discourse-level analysis to extract biological relations that might potentially lead to serendipitous discovery. We found that biological relations quite often transcend clausal sentences and hence demand mechanisms to connect information across such boundaries. Using drug-disease relation discovery as an example, the description of 23,268 potential drug-disease relations across sentence boundaries against the 14,451 relations within sentences indicates the importance of a discourse-level analysis requirement to extract these relations from biomedical literature.

Across all experiments we observed a consistent trend that drug-disease relations that occurred across sentences outnumbered the ones within sentences. Our evaluation against a curated resource such as CTD also reinforces this fact. Results from two experiments need attention 1) Frequency of drug-disease relation co-occurrence at sentence level against the discourse level (Figure 3) was skewed towards the latter than the former. 2) The time lag analysis where the discourse-level analysis would have significantly reduced the time lag between the scientific reporting of causal pairs (drug-gene/gene-disease) and drug-disease relational pairs. This shows the need for advance discourse-level analysis approaches to extract information from literature in time and its ability to hasten the pace of discovery.

Another significant observation is the amount of potential false-positive drug-disease relations identified through literature mining. We observed a substantial reduction in the number of drug-disease relations when we compared the literature based drug-disease pairs with that of curated resource CTD. The reduction among the discourse-level pairs is far greater (3.3 times) when compared to the ones from the sentence level (2.5 times). As a note of caution discourse-level analysis might potentially extract more false positives, provided one does not explore sophisticated linguistic/statistical approaches to handle them.

As a final note we would like to mention that the way we generate the seed pairs of drug-disease relations in this study, coupled with our timeline based restriction analysis reinforces that the goal of our study is centered on LBD of novel relations. However we also understand the limitations of this study. Simple sentence-level or document-level co-occurrence information does not imply any biological relation between entities. As pointed out earlier, simple discourse-level analysis has the danger of extracting more false positives. In certain tangential areas of research such as drug repurposing certain techniques [28], [29] were explored to reduce the false positives.

Conclusion

In this paper, we investigated the extent of the need of discourse-level analysis for drug-disease relation extraction from biomedical literature. We used Semantic Medline to extract LBDs and then, based on co-occurrence analysis, we collected any evidence of the discoveries in Medline abstracts. We categorized the evidence into two categories, sentence level and discourse level. From subsequent analysis we infer that there is a potential to miss more than 70% of drug-disease relations when we extract information from the sentence level. This clearly demonstrates the need for deeper discourse-level analysis, which may translate to significant improvement in the state of the art of NLP techniques.

In the near future, we plan to explore much more sophisticated NLP approaches, a significant departure from the co-occurrence based extraction which may lead to significant improvement in the state of the art.

Acknowledgments

This study was made possible by the National Institutes of Health grant R01 GM102282-01.

References

[1] D. Marcu and A. Echihabi, *An Unsupervised Approach to Recognizing Discourse Relations*. 2002.

[2] P. W.-H. Johanna D Moore, "Discourse in Computational Linguistics and Artificial Intelligence."

[3] H. Kilicoglu, D. Shin, M. Fiszman, G. Rosemblat, and T. C. Rindflesch, "SemMedDB: a PubMed-scale repository of biomedical semantic predications," *Bioinformatics*, vol. 28, no. 23, pp. 3158–3160, Dec. 2012.

[4] R. R. V. Goulart, V. L. Strube de Lima, and C. C. Xavier, "A systematic review of named entity recognition in biomedical texts," *J Braz Comput Soc*, vol. 17, no. 2, pp. 103–116, Mar. 2011.

[5] O. Ghiasvand, "Disease Name Extraction from Clinical Text Using Conditional Random Fields," *Theses and Dissertations*, May 2014.

[6] N. Bach and S. Badaskar, *A Review of Relation Extraction*. .

[7] M. Sun and J. Y. Chai, "Discourse processing for context question answering based on linguistic knowledge," *Knowledge-Based Systems*, vol. 20, no. 6, pp. 511–526, Aug. 2007.

[8] S. Kim, R. Bracewell, and K. Wallace, "From discourse analysis to answering design questions",Proceedings," in *of the Workshop on the Application of Language and Semantic Technologies to support Knowledge Management Processes*, 2004, pp. 43–49.

[9] L. B. Suzan Verberne, "Evaluating discourse-based answer extraction for why-question answering," 2007.

[10] H. Prendinger, P. Piwek, and M. Ishizuka, "Automatic Generation of Multi-Modal Dialogue from Text Based on Discourse Structure Analysis," in *International Conference on Semantic Computing, 2007. ICSC 2007*, 2007, pp. 27–36.

[11] M. Zahedi and M. Kahani, "SREC: Discourse-level semantic relation extraction from text," *Neural Comput & Applic*, vol. 23, no. 6, pp. 1573–1582, Nov. 2013.

[12] L. Qian and G. Zhou, "Tree kernel-based protein–protein interaction extraction from biomedical literature," *Journal of Biomedical Informatics*, vol. 45, no. 3, pp. 535–543, Jun. 2012.

[13] C. Quan, M. Wang, and F. Ren, "An Unsupervised Text Mining Method for Relation Extraction from Biomedical Literature," *PLoS ONE*, vol. 9, no. 7, p. e102039, Jul. 2014.

[14] M. Bundschus, M. Dejori, M. Stetter, V. Tresp, and H.-P. Kriegel, "Extraction of semantic biomedical relations from text using conditional random fields," *BMC Bioinformatics*, vol. 9, no. 1, p. 207, Apr. 2008.

[15] M. Rastegar-Mojarad, "Extraction and Classification of Drug-Drug Interaction from Biomedical Text Using a Two-Stage Classifier," *Theses and Dissertations*, Dec. 2013.

[16] S. Mehrabi, A. Krishnan, E. Tinsley, J. Sligh, N. Crohn, H. Bush, J. Depasquale, J. Bandos, and M. Palakal, "Event Causality Identification Using Conditional Random Field in Geriatric Care Domain," in *2013 12th International Conference on Machine Learning and Applications (ICMLA)*, 2013, vol. 1, pp. 339–343.

[17] D. R. Swanson, "Migraine and magnesium: eleven neglected connections," *Perspect. Biol. Med.*, vol. 31, no. 4, pp. 526–557, 1988.

[18] M. Weeber, H. Klein, L. T. W, J. Berg, and D. R. S. Has, "Using concepts in literature-based discovery: Simulating Swanson's Raynaud-fish oil and migrainemagnesium discoveries," *J. Am. Soc. Inf. Sci. Tech*, pp. 548–557, 2001.

[19] D. Hristovski, B. Peterlin, and S. Dzeroski, "Literature-based Discovery Support System and Its Application to Disease Gene Identification," *Proc AMIA Symp*, p. 928, 2001.

[20] M. Yetisgen-Yildiz and W. Pratt, "A new evaluation methodology for literature-based discovery systems," *J Biomed Inform*, vol. 42, no. 4, pp. 633–643, Aug. 2009.

[21] C. Andronis, A. Sharma, V. Virvilis, S. Deftereos, and A. Persidis, "Literature mining, ontologies and information visualization for drug repurposing," *Brief. Bioinformatics*, vol. 12, no. 4, pp. 357–368, Jul. 2011.

[22] D. Hristovski, C. Friedman, T. C. Rindflesch, and B. Peterlin, "Exploiting Semantic Relations for Literature-Based Discovery," *AMIA Annu Symp Proc*, vol. 2006, pp. 349–353, 2006.

[23] T. C. Rindflesch and M. Fiszman, "The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text," *J Biomed Inform*, vol. 36, no. 6, pp. 462–477, Dec. 2003.

[24] M. Rastegar-Mojarad, D. Li, and H. Liu, "Operationalizing Semantic Medline for meeting the information needs at point of care," presented at the AMIA Clinical Research Informatics Summit, 2015.

[25] S. Moosavinasab, M. Rastegar-Mojarad, hongfang Liu, and S. Jonnalagadda, "Towards Transforming Expert-based Content to Evidence-based Content," presented at the AMIA Clinical Research Informatics Summit, 2014.

[26] A. P. Davis, C. J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B. L. King, T. C. Wiegers, and C. J. Mattingly, "The Comparative Toxicogenomics Database's 10th year anniversary: update 2015," *Nucleic Acids Res.*, Oct. 2014.

[27] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart, "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs," *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D1035–1041, Jan. 2011.

[28] H. Xu, M. C. Aldrich, Q. Chen, H. Liu, N. B. Peterson, Q. Dai, M. Levy, A. Shah, X. Han, X. Ruan, M. Jiang, Y. Li, J. S. Julien, J. Warner, C. Friedman, D. M. Roden, and J. C. Denny, "Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality," *J Am Med Inform Assoc*, Jul. 2014.

[29] M. Rastegar-Mojarad, Z. Ye, J. M. Kolesar, S. J. Hebbring, and S. M. Lin, "Opportunities for drug repositioning from phenome-wide association studies," *Nat Biotech*, vol. 33, no. 4, pp. 342–345, Apr. 2015.

Address for correspondence

The correspondence for this study is Majid Rastegar-Mojarad. His email address is Mojarad.Majid@mayo.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-544

InfoRoute: the CISMeF Context-specific Search Algorithm

Tayeb Merabti[a], Romain Lelong[a], Stefan Darmoni[a,b]

[a]*CISMeF & TIBS, LITIS EA 4108, Rouen University Hospital, Rouen, France*
[b] *INSERM, U1142, LIMICS, F-75006, Paris, France*

Abstract

Objective*. The aim of this paper was to present a practical InfoRoute algorithm and applications developed by CISMeF to perform a contextual information retrieval across multiple medical websites in different health domains.* ***Methods.*** *The algorithm was developed to treat multiple types of queries: natural, Boolean and advanced. The algorithm also generates multiple types of queries: Boolean query, PubMed query or Advanced query. Each query can be extended via an inter alignments relationship from UMLS and HeTOP portal.* ***Results.*** *A web service and two web applications have been developed based on the InfoRoute algorithm to generate links-query across multiple websites, i.e.: "PubMed" or "ClinicalTrials.org".* ***Conclusion.*** *The InfoRoute algorithm is a useful tool to perform contextual information retrieval across multiple medical websites in both English and French.*

Keywords:
Algorithms; Information Storage and Retrieval; Medical Informatics Applications.

Introduction

The Internet and in particular the Web has become an extensive health information and knowledge repository. As described in the survey published by Podichetty et al. [1], a majority (72%) of physicians affirmed that they used the Internet on a regular basis for medical research and 51% of them confirmed that the Internet influenced their healthcare practice. However, as mentioned by De Leo *et al.* [2] the vast majority of physicians (92%) indicated that they access a target site rather than utilize a search engine to gather medical information. A comparison study by Freeman *et al.* [3] concluded that PubMed appears to be more specific than Google Scholar; which was the same conclusion reported by Anders *et al.*[4] "PubMed is more practical to conduct efficient and valid searches on clinical topics than Google Scholar".

With the explosion of available health information and the number of biomedical websites in different medical fields including: literature databases, clinical trials databases, and drug databases; knowledge and information retrieval (IR) have become more complex and more time consuming. In the medical field some problems have obstructed the IR [5],[6] including the following:

- **Inadequate expression of information needs:** many information needs cannot be appropriately expressed: query syntax, medical terminologies or languages.
- **Lack of the relevant information:** Due to the inadequate terms used to search information, as medical databases and IR systems are indexed according to specific health terminologies which are different: lexically (different terms referred to the same medical concept) and structurally (different relationships between same concepts).

The aim of this work is to propose an algorithm named "InfoRoute" to perform a contextual and cross-lingual information retrieval across multiple websites in different health fields. This algorithm is devoted to French-speaking health professionals, who can read English but who prefer to perform queries in their native language. InfoRoute builds and generates various IR query on websites such as: PubMed, ClinicalTrials.org. Besides the algorithm, a web service and multiple web applications have been developed to help physicians and students to express the adequate IR query including medical terms in English and French based on more than 56 health terminologies.

Materials

The UMLS Metathesaurus

The UMLS Metathesaurus [7], developed by the US National Library of Medicine (NLM®), integrates 2,930,638 concepts in its 2014 release from 168 biomedical vocabularies. In the "MRCONSO"[1] table, which lists all UMLS concepts, only four terminologies, of all UMLS terminologies, are included with their French version in the UMLS Metathesaurus: the MeSH thesaurus, the World Health Organization Adverse Reaction Terminology (WHO-ART), the WHO International Classification of Primary Care (ICPC2), and the Medical Dictionary of Regulatory Activities (MedDRA). However, five (5) Biomedical Terminologies and Ontologies (BMTO) that have an existing official French version, are included in the UMLS but without their French version: the International Statistical Classification of Diseases (ICD10), the Systematized NOmenclature of MEDicine (SNOMED Int), Logical Observation Identifiers Names and Codes (LOINC), the International Classification of Functioning, Disability and Health (WHO-ICF) for handicap and the International Classification for Nursing Practice (ICNP). Furthermore, the CISMeF (French acronym for "catalogue and index of online medical resources in French") team has partially translated the BMTO included in the UMLS only in English: 24,563 synonyms and 689 ambiguous acronyms of the MeSH Descriptors, 163 synonyms of the MeSH Qualifiers, 20,887 MeSH Supplementary Concepts, and, 847 MEDLINEPlus terms and 12,700 FMA terms.

[1] The name of the concepts table in the UMLS metathesaurus

The Health Multiple Terminologies and Ontologies Portal (HeTOP)

A generic meta-model was designed in order to fit all 56 terminologies into one global structure.

The HeTOP[8] is connected to this meta-model to search concepts from all the health terminologies available in French (or in English and translated into French) included in this portal and, to browse it dynamically. This allows to:

a) Manual or automatic indexing of resources for the catalogue;
b) Retrieval of resources;
c) Teaching or performing audits in terminology management.

Some terminologies and classifications are included in the UMLS Metathesaurus (N=17) but the majority are not (N=39), i.e. the Human Rare Diseases Ontology (HRDO). Currently, HeTOP integrates 1,743,772 concepts in English, 1,031,230 in French, and 9,255,438 relationships.

The HeTOP portal integrates multiple terminological relationships and they can be classified as:

1. **Intra Terminological relationships:** terms linked with this type of relationships are in the same terminologies. For example, the MeSH term "suicide" is linked according to the Intra MeSH relationships **"See also"** with the MeSH term "suicide, attempted".
2. **Inter Terminological relationships:** these relationships link terms from different terminologies. In HeTOP, four (4) main inter-relationships were integrated and/or created (see *Table 1* for examples):
   a. **UMLS alignment ($UMLS_{alignment}$) :** This conceptual relationship described previously by Merabti *et al.*[9]. Two terms in HeTOP are linked under this relationship if they share the same UMLS concept in the Metathesaurus.
   b. **Manual inter-alignment ($CISMeF_{manual}$):** This relationship is added manually in HeTOP by CISMeF physicians or terminological specialist to link between two similar terms described in the same medical concept.
   c. **Exact inter-alignment ($CISMeF_{exact}$):** This relationship is obtained automatically using the lexical approach previously described by Merabti *et al.*[9]. Two terms are linked using $CISMeF_{exact}$ if they are lexically similar at the preferred term or at the synonym level, in French or English.
   d. **Exact supervised inter-alignment ($CISMeF_{supervised}$):** There are several $CISMeF_{exact}$ relationships validated by CISMeF physicians and tagged as $CISMeF_{supervised}$. In terms of efficiency this relationship is equivalent to $CISMeF_{manual}$ relationship.

Methods

Websites categorization

A total of twelve (12) categories were created to classify multiple medical websites (see Table 2). The selection of websites was performed by librarians according to the needs of health professionals, students and patients. Moreover, some websites were considered as the most important: e.g. PubMed, US Clinical Trials and EU Clinical Trials. Therefore, specific functions for these websites were developed. For example, in the following sections we described a PubMed query which is a specific querying on PubMed. The websites selected are important since it conditions the input and output of the algorithm developed.

Table 1 –examples from each 4 inters relationships integrated and their number in HeTOP (2014 released)

	Source Term (Terminology)	Target Term (Terminology)	Number of relations in HeTOP
$UMLS_{alignment}$	*Myocardial Infarction (MeSH)*	*Myocardial infarction, NOS (SNOMED Int)*	*644,982*
$CISMeF_{manual}$	*Riedel thyroiditis(HRDO)*	*Riedel's thyroiditis (MedDRA)*	*41,673*
$CISMeF_{exact}$	*appetite stimulants (ATC)*	Appetite stimulated (WHOART)	*653,709*
$CISMeF_{Supervised}$	*gonadotropin-releasing hormone (MeSH)*	*Luteotropin-releasing factor (FMA)*	*251,995*

InfoRoute Algorithm

InfoRoute is defined as an algorithm that automatically generates multiple search queries across many medical websites and exploiting the entire range of French and English Medical terminologies included in the HeTOP and some terminologies from the UMLS Metathesaurus.

The input

The InfoRoute algorithm takes in input three (3) kinds of queries:

1. **Natural Input Query (NI_{Query}):** this type of query assumes that the input text is composed of multiple terms in English and French in the natural language without using Boolean or restricted operators.
2. **Boolean Input Query (BI_{Query}):** query is based on the CISMeF Boolean query syntax shown in an equation 1. Where:
   a. **RO:** corresponds to the restricted operators (ROs)[2] as defined by the CISMeF such as "mr" for reserved words, "ti" for title or "tc" as all fields. These ROs can be useful to accurately translate the query to PubMed.
   b. **TERMINO:** the terminologies assigned to each term.

In addition to these ROs, the following terms can be connected using Boolean operators: AND, OR and NOT.

For example, the query **"asthma.mr[TER_MSH] AND child.tc"** corresponds to the query "search resources indexed by asthma as a reserved MeSH term and with the term child in all fields (title, abstract, etc.).

3. **Advanced Input Query (AI_{Query}):** this type of query is composed of at least one natural or Boolean query

$$(\text{Term}(.\text{RO}([\text{TERMINO}])^*)^+((\text{AND}|\text{OR}|\text{NOT})(\text{Term}(.\text{RO}([\text{TERMINO}])^*)^+)^* \quad (1)$$

and some options such as: age, country, gender. These options are used to translate the query to clinical trial websites: ClinicalTrials, Clinical Trials Register.

[2] Restricted operators used in the CISMeF search engine: http://doccismef.chu-rouen.fr/aides/aidedcacronyme.html [Nov 2014]

In addition to the query, the algorithm is a terminology depending, therefore, according to the terminology assigned in an input, and the results change whether query terms are included or not.

The Mutli-Terminological Automatic Indexing Query (MTAIQ)

This part of algorithm is very important, since in this query stage the input is indexed to extract the most similar medical terms in the query. The MTAIQ is a multi-terminological-bilingual indexer [9]. It uses the HeTOP databases and natural language processing tools [9] to analyze and normalize the query. In the case of the natural query, MTAIQ maps the query to the most similar term(s) depending on terminologies assigned in the input. For example, if the query is **"childhood asthma"** then MTAIQ detects "asthma in children" term, since the "childhood asthma" is a synonym of the MEDLINEPlus term "asthma in children". Nevertheless, if only the MeSH terminology is assigned then the most similar term detected will be "**asthma**", since "childhood asthma" and childhood are not MeSH terms.

The query Translation

In order to query all the websites, the InfoRoute algorithm performed multiple query translation from the one of the three queries types to at least four (4) possible kinds of queries:

1. **Boolean Output Query (BO_{Query}):** the query generated is composed by terms indexed using the MTAIQ and separated with Boolean operators. For example, if the query is **"furlong syndrome"** and the terminology selected is MeSH then the query generated will be "(**Furlong syndrome) OR (marfanoid disorder with craniosynostosis, type 2".** In this example the entry and the input languages were in English but it is possible to translate the result into the corresponding French Boolean query: **(syndrome de furlong) OR (furlong) OR (craniosynostose marfanoide)**".
2. **Extended Boolean Output Query (EBO_{Query}):** the Boolean query generated can be extended using the inter terminology relationships: $UMLS_{alignement}$, $CISMeF_{manual}$ and $CISMeF_{supervised}$ relationships described above. The aim was to obtain a new query with additional synonyms not included in the original user query. In this case, synonyms are all terms preferred or synonyms related to the query terms in the same language and according to the medical terminologies assigned.
3. **PubMed Query ($PubMed_{Query}$):**

Most of the research efforts were concentrated on PubMed queries because PubMed is of utmost importance for health professionals, students and patients. Some improvements have already been performed by CISMeF on PubMed query (vs. the PubMed query by default) since 2009[11],[12]. The first study was testing the added value of MeSH synonyms (MeSH Entry Terms) [11] and the second study was testing the added value of UMLS synonyms (same CUI) [12]. The PubMed query is a Boolean query combined MeSH terms (preferred and synonyms) and search options from PubMed: [MH]: Main Headings, [TI]: Title or [TW] for text words. This query was generated from natural or Boolean query. In addition to the simple PubMed query, in CISMeF two additional PubMed extended queries and one PubMed manually build query were developed:

a. **PubMed UMLS-Extended Query ($PubMedUmlsE_{Query}$):** As the EBO_{Query} query and as described in [12]. The PubMed query was extended by adding synonyms from UMLS representing terms which are in relation with the MeSH terms in the query and translated into English.

b. **PubMed Inter Extended Query ($PubMedInterExt_{Query}$):** Like the $PubMedUmlsE_{Query}$, the PubMed query was extended by adding synonyms which are in related to the MeSH terms according to two (2) inter relationships: $CISMeF_{manual}$ and $CISMeF_{supervised}$. Furthermore, for some BMTO included in HeTOP a PubMed generator query was developed based on these relationships. This generator extracted only the MeSH terms (MeSH descriptor and MeSH Supplementary concepts) for "no MeSH terms" which were related to it and build automatically a PubMed query using these MeSH terms. For example, for the HRDO term "Marfan syndrome" will be extended to the MeSH term "marfan syndrome" and all its synonyms.

c. **PubMed Manual Query ($PubMedManual_{Query}$):** In some cases where the automatic generated query was not effective (for the drug terms as an example), a manually PubMed query for each term was created in order to be more accurate. In a previous study[13], the "ATCtoPubMed" application to access PubMed via any Anatomical Therapeutic Chemical Classification (ATC) code was described. For each ATC code, a predefined query was created and could be entered on PubMed.

4. **Advanced Clinical Query ($Advanced_{Query}$):** This query is based on the AI_{Query} since it generates a "Boolean" query combined with the AI_{Query} Options selected. Currently, options have been selected based on two clinical trials websites: http://clinicaltrials.gov (US clinical trials) and http://www.clinicaltrialsregister.eu (EU clinical trials). Moreover, one name options were used as input and the $Advanced_{Query}$ mapped options in input to the adequate name option in the query search for each clinical websites. For example, the option "SEXE=female" will be mapped to: "gender=female-only" in EU clinical trials and to "gndr=Female" in the US clinical trials.

Table 2– Principals Websites classified by categories and language.

Categories	Website	
	English	**French**
Clinical Trials	1. ClinicalTrials: http://clinicaltrials.gov 2. Clinical Trials Register: http://www.clinicaltrialsregister.eu	1. ANSM: http://ansm.sante.fr
Drugs	1. MedlinePlus: http://www.nlm.nih.gov/medlineplus/ 2. Drug Information Portal: http://druginfo.nlm.nih.gov/drugportal	1. Portail D'information sur les medicaments: http://doccismef.chu-rouen.fr/servlets/PIM http://www.has-sante.fr/
Information For Patients	1. NIH Senior Health: http://nihseniorhealth.gov/ 2. MedlinePlus: http://www.nlm.nih.gov/medlineplus/	1. CISMeF Patient: http://doccismef.chu-rouen.fr/dc/#env=pa
Public Health		Banque de données en santé publique http://www.bdsp.ehesp.fr/

Rare Diseases	1. Gentics Home reference: http://ghr.nlm.nih.gov/ 2. NCBI OMIM: http://www.ncbi.nlm.nih.gov/omim?	1. Orphanet: http://www.orpha.net
Search Engine	1. PubMed: http://www.ncbi.nlm.nih.gov/pubmed 2. NLM Gateway: http://gateway.nlm.nih.gov/gw/Cmd/home.jsp	. CISMeF: http://doccismef.chu-rouen.fr/dc/
Students	1. PubMed: http://www.ncbi.nlm.nih.gov/pubmed	1. DocUMVF: http://doccismef.chu-rouen.fr/dc/#env=umvf

Table 3– Examples of queries types generated by InfoRoute Algorithm.

Query Search Type	Query	Query Terminology	InfoRoute query
BO_{Query}	abnormal platelets	*MedDRA*	(abnormal platelets) OR (platelets abnormal) OR (thrombocytes abnormal (nos))
EBO_{Query}	abnormal platelets	*MedDRA*	(abnormal platelets) OR (platelets abnormal) OR (thrombocytes abnormal (nos)) OR (platelet abnormalities) OR (abnormal platelet) OR (atypical platelet) OR (glanzman s disease) OR (hereditary thrombasthenia) OR (platelet changes) OR (thromboasthenia) OR (thrombasthenia)
$PubMed_{Query}$	alcohols	*MeSH*	(("alcohols"[MH] OR ("alcohols"[TW]))
$PubMedUmlsExt_{Query}$	alcohols	*MeSH*	(("alcohols"[MH] OR ("alcohols"[TW] OR "alcohol, nos"[TW]))
$PubMedInterExt_{Query}$	alcohols	*MeSH*	(("alcohols"[MH] OR ("alcohols"[TW] OR "Ethanol"[TW] OR "Alcohol"[TW] OR "drinking"[TW] OR "alcohol consumption"[TW] OR "alcohol, nos"[TW]))
$PubMedManual_{Query}$	allergens	*ATC*	"desensitization, immunologic"[MH] AND "allergens"[MH]"

Results

The InfoRoute Web Service

The web service can be used by third-party web services or web-based applications. For example, The HeTOP portal used the InfoRoute to access PubMed from any BMTO's terms. In the CISMeF catalog, a specific website link which can be refreshed after any CISMeF query is integrated. The website category corresponds to the "search engine category" from the twelve (12) categories introduced above: PubMed, Intute, NCBI GQuery, etc.

InfoRoute Web Applications

Two web applications were devloped based on InfoRoute:

1. The InfoRoute General Website Tool

The application at http://inforoute.chu-rouen.fr/ir, visualises the input and the search results of the InfoRoute algorithm described in this paper. The input query allows the end-user to enter a term or an expression in French or English. All terminologies are assigned to the query by default. However, it is possible to specify the terminology(ies) input. Term or expression can be expressed using NI_{Query} or BI_{Query} described above. Search results can be obtained from all the twelve (12) category websites. Each category will be displayed separately and websites in each category will be separated according to their languages (French or English).

2. The InfoRoute Clinical Trials Access

The application at http://inforoute.chu-rouen.fr/irec, can be considered as a subtype of the InfoRoute general website tool described above. Tha aim of this application is to perform an effective access especially in French but also in English for two clinical trial websites: http://clinicaltrials.gov (US clinical trials) and http://www.clinicaltrialsregister.eu (EU clinical trials). Additional options will be added to this application compared to the general application based on AI_{Query} to generate and $Advanced_{Query}$ described in the query translation section.

Discussion

The aim of this study was to present the InfoRoute algorithm, which uses multiple BMTO and multiple relationships to generate IR health queries from natural or Boolean queries in French or English. The InfoRoute algorithm permits to answer the two questions presented in the introduction:

- **Inadequate expression of information needs:** the use of multiple kinds of queries (natural, Boolean and advanced), helps users to express their needs in multiple forms in two languages. Furthermore, the algorithm permits the user to express their queries according to multiple BMTO directly from specific biomedical terms or indirectly when applied the MTAIQ to extract biomedical terms.
- **Lack of the relevant information:** As described for PubMed and some clinical websites; the algorithm has been improved to retrieve the more accurate information for these specific websites. For example, in Thirion *et al.*[11] the optimized query is significantly more precise than the current PubMed query (54.5% vs. 27%). In [12], the expansion of the PubMed query on UMLS synonyms increase recall and proportion of queries retrieving. However, the expansion of the user query according to multiple synonyms can decrease precision because the description of the same medical concept differ between BMTO. Besides UMLS synonyms, the $PubMedInterExt_{Query}$ which based on synonyms terms expanded using $CISMeF_{manual}$ and $CISMeF_{supervised}$ relationships will be evaluated in further work. Further studies will be performed to exploit hierarchical inter relationships to expand results query.

The InfoRoute algorithm is the first algorithm developed which used French terms and/or French BMTO to generate and reformulate IR queries according to multiple health websites such as: PubMed or ClinicalTrials. The use of two languages is due to the bilingual BMTO included in HeTOP and thanks to multiple automatic and manual translations performed on multiple BMTO such as: MEDLINEPlus, MeSH SC, FMA, etc. Currenlty InfoRoute is integrated in the two main CISMeF tools: CISMeF search engine (http://doccismef.chu-rouen.fr/dc/) and HeTOP (http://www.hetop.eu/hetop/) to retrieve health infrormations from many sources. The next step is to integrate InfoRoute in real clinical information systems and develop a "French InfoButtons" as described in [14].

Conclusion

To conclude, the InfoRoute algorithm is a useful tool to perform contextual information retrieval across multiple medical websites in both English and French.

Acknowledgments

This work was partially granted by: TerSan project (ANR TecSan program n°ANR-11-TECS-019-03). The ReLySe project (REfractory LYmphoma Sequencing) Transla 2013. The Authors are grateful to Richard Medeiros, for editing assistance.

References

[1] Podichetty V. Booher J., Whitfield M. Biscup R. A. Assessment of internet use and effects among healthcare professionals: a cross sectional survey. Postgrad. Med. J. 2006; 2:2744-9.

[2] De Leo G, LeRouge C, Ceriani C, Niederman F. Websites most frequently used by physician for gathering medical information. AMIA Annual Symposium Proceedings, 2006; pp. 206.

[3] Freeman MK, Lauderdale SA, Kendrach MG, Woolley TW. Google Scholar versus PubMed in locating primary literature to answer drug-related questions. Ann Pharmacother 2009;43:478.

[4] Anders ME, Evans DP. Comparison of PubMed and Google Scholar literature searches. Respiratory care. 2010, 55:578-83.

[5] Braun L. Pro-active medical information retrieval. PhD Thesis, 2009.

[6] ter Hofstede AHM, Proper HA, van der Weide ThP. Query Formulation as an Information Retrieval Problem. The Computer Journal 1996; 39;pp. 255-274.

[7] Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med. 1993;32(4):281–291.

[8] Grosjean J, Merabti T, Dahamna B, Kergouraly I, Thirion B, Soualmia L, Darmoni, SJ. Health Multi-Terminology Portal: a semantic added-value for patient safety. In: PSIP Workshop; 2011. p. 129–138.

[9] Merabti T, Soualmia LF, Grosjean J, Joubert M & Darmoni SJ. Aligning Biomedical Terminologies in French: Towards Semantic Interoperability in Medical Applications. In Book: Medical Informatics, pp. 41-68, InTech, 2012.

[10] Pereira S.: Multi-terminology indexing of concepts in health. [Indexation multi-terminologique de concepts en santé]. PhD Thesis, University of Rouen, Normandy,France.

[11] Darmoni SJ, Sakji S, Pereira S, Merabti T, Prieur E, Joubert M, Thirion B. Multiple terminologies in an health portal: automatic indexing and information retrieval. Artificial Intelligence in Medicine, Lecture Notes in Computer Science; 2009:255-259.

[12]Thirion B, Robu I, Darmoni, SJ. Optimization of the PubMed Automatic Term Mapping. Stud Health Technol Inform.2009;150: 238-242.

[13]Griffon N, Chebil W, Rollin L, Kerdelhue G, Thirion B, Gehanno JF, Darmoni SJ. Performance evaluation of Unified Medical Language System®'s synonyms expansion to query PubMed. BMC Medical Informatics and Decision Making. 2012;12(1):12.

[14]Merabti T, Abdoune H, Letord C, Sakji S, Joubert M, Darmoni SJ. Mapping the ATC classification to the UMLS Metathesaurus: some pragmatic applications. Patient Safety Informatics - Adverse Drug Events, Human Factors and IT Tools for Patient Medication Safety. Studies in Health Technology and Informatics, Volume 166, Pages 206-213, 2011.

[15]Darmoni SJ, Pereira S, Névéol A, Massari P, Dahamna B, Letrod C, Kerdeluhé G, Piot J, Derville A, Thirion B. French Infobutton: an academic and…business perspective. AMIA Annual Symposium Proceedings;pp. 920,2008.

Address for correspondence

Tayeb Merabti , CISMeF & TIBS, LITIS EA 4108, 1 rue de Germont 76031 Rouen Cedex, France, Tayeb.merabti@chu-rouen.fr.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-549

A Baseline Patient Model to Support Testing of Medical Cyber-Physical Systems

Lenardo C. Silva[a], Mirko Perkusich[b], Hyggo O. Almeida[a], Angelo Perkusich[c], Mateus A. M. Lima[c], Kyller C. Gorgônio[a]

[a] *Embedded Systems and Pervasive Comp. Lab, Federal University of Campina Grande, Campina Grande, Brazil*
[b] *Federal Institute of Paraíba, Monteiro, Brazil*
[c] *Electrical Engineering Department, Federal University of Campina Grande, Campina Grande, Brazil*

Abstract

Medical Cyber-Physical Systems (MCPS) are currently a trending topic of research. The main challenges are related to the integration and interoperability of connected medical devices, patient safety, physiologic closed-loop control, and the verification and validation of these systems. In this paper, we focus on patient safety and MCPS validation. We present a formal patient model to be used in health care systems validation without jeopardizing the patient's health. To determine the basic patient conditions, our model considers the four main vital signs: heart rate, respiratory rate, blood pressure and body temperature. To generate the vital signs we used regression models based on statistical analysis of a clinical database. Our solution should be used as a starting point for a behavioral patient model and adapted to specific clinical scenarios. We present the modeling process of the baseline patient model and show its evaluation. The conception process may be used to build different patient models. The results show the feasibility of the proposed model as an alternative to the immediate need for clinical trials to test these medical systems.

Keywords:

Baseline Patient Model; Statistical Analysis; Simulation; Testing; Medical CPS.

Introduction

In the health care domain, traditional clinical scenarios are seen as *closed-loop systems* in which the caregivers are the *controllers*, medical devices act as *sensors* and *actuators*, and patients are the "*physical*" *plants*. *Medical Cyber-Physical Systems* (MCPS) modify this vision by introducing additional computational entities that help the caregiver control the "*plant*", i.e., decision support [1].

Due to the insufficient understanding of human body dynamics in response to any treatment, MCPS development is more complex than traditional systems [1]. Since patient safety is the main concern of an MCPS, they must be carefully tested before being released to the market. However, to allow a MCPS to be directly tested with real patients, it is necessary to have an alternative solution for developers to identify critical errors and prevent serious harm to test patients. Therefore, a pre-validation step must be added to the validation process to anticipate the identification of these errors.

Nowadays, MCPS designers, more specifically the manufacturers of medical devices, use documents provided by regulatory agencies related to health, such as the FDA[1]. They specify standards for safety and performance to minimize design risks and to ensure the effectiveness of products.

The importance of having a formal patient model is the capability of generating relevant test cases to validate an MCPS. This process aims to examine if the therapies provided by these systems are adequate and ensure patient safety by adapting their behavior given the patient's current state [2].

The concept of a *virtual patient* described by Agur [3] is related to a complex set of mathematical models and a set of parameters that represent the dynamics of biological, pharmacological, and pathological processes in the body of a patient under medication. In the literature, we can find some related work about computational modeling of patients. For instance, Jiang et al. [2] provide an environment for closed-loop testing, whose patient model, specifically a formal model of the human heart, is the control center of a cardiac pacemaker system. The goal is to evaluate the device's operation safety and effectiveness based on the patient condition.

Other recent research is presented by Van Heusden et al. [4], which proposes an artificial pancreas model for patients with type 1 diabetes mellitus based on control theory. Its goal is to improve glucose control in such patients. One of its advantages is that it incorporates the patient's medication (insulin) reaction.

Lastly, Khan et al. [5] present a glucose control system to prevent hypoglycemia as a consequence of medication administration. In this work, the patient model (i.e., an artificial pancreas) establishes a relationship between certain vital signs such as heart rate and skin impedance in the blood glucose level control.

The aforementioned patient models either focus on variables of interest for specific clinical scenarios or neglect the relationship between the four human being's vital signs: heart and respiratory rates, blood pressure, and body temperature. Therefore, the absence of these aspects contradicts the actual behavior of human beings. So, these models have limited applicability to other scenarios.

In this paper, we propose a formal patient model to be used as a basis for developers to validate MCPS solutions without risking compromising patients' health if the system fails. Our model provides parameters to define patient profiles, and correlates the main vital signs to simulate the patient's health condition. The vital signs are represented by statistical regression models which were integrated to provide the patient's model dynamics.

[1] United States Federal Food and Drug and Administration - http://www.fda.gov/

The key feature of the baseline patient model is the ability to be adapted to various specific clinical scenarios, promoting reuse. The main contribution of this paper is to present the conception process used to develop the baseline patient model of the model-based architecture proposed by Silva et al. [6].

Materials and Methods

The vital signs and physiological parameters provided by our patient model are based on thresholds. They are generated by regression models. We established the parameters that defined the thresholds in accordance with Clinical Guidelines, such as [7-9]. These clinical guidelines also were used to help identify the probable regression model predictor variables for each vital sign considered in our baseline patient model. Also, we applied some rules on the real patient data obtained from the *MIMIC II Clinical Database v2.6*, whose access was authorized by the PhysioNet.org [10]. This data set contains clinical data from Intensive Care Unit (ICU) patients and the rules were applied to characterize the population of interest for study. The baseline patient model conception process consists of: (1) Statistical analysis and (2) Computational modeling.

Statistical Analysis

We used the same database to characterize a population of individuals. We applied inferential statistical techniques such as a *Generalized Linear Model* (GLM) to obtain regression models. These models are used to predict the observed signs to be incorporated in the baseline patient model.

In the GLM, is assumed that the response variable follows the exponential family distribution, and the predicted values are calculated from a link function [11].

The population contained 38,141 observations from 2,245 patients, in which approximately 37.4% were female and 62.6% male. From this population, we obtained a sample with one observation of each patient in the moment they were admitted to the ICU. We present a summary of the process to characterize the population of interest and sampling procedure in Figure 1, whose sample size represents only approximately 0.001% of total population. Table 1 presents descriptive statistics about the data set extracted from this process.

Table 1 – Statistics of the Population of Interest

Variable	Mean	Std. Dev.	CV [a]	Min.	Max.
hr	86.863	14.473	0.167	42.00	150.00
sbp	115.447	20.403	0.177	62.00	205.00
rr	17.552	5.619	0.320	8.00	38.00
pt	36.926	0.879	0.024	31.70	41.44
gl	129.631	30.444	0.235	47.00	188.00
weight	83.727	20.826	0.249	33.00	200.00
height	169.781	10.427	0.061	124.50	231.10

[a] Coefficient of Variation; Std. Dev. = Standard Deviation, Min = minimum, Max. = Maximum.

We selected eight variables for the statistical analysis: *gender*, *weight* and *height* as demographic variables; heart rate (*hr*), respiratory rate (*rr*), systolic blood pressure (*sbp*), and body temperature (*pt*), as vital signs; and blood glucose level (*gl*), as a physiological parameter. It is worth mentioning that a variable for diastolic blood pressure was not used in this analysis because it depends on *sbp* variable. The strong correlation between these variables may cause a multicollinearity problem [12].

Given that the population of interest was defined and one sample was obtained, we extracted four regression models, one for each vital sign in study. The regression models were evaluated using the following statistical measures: (a) determination coefficient (R^2); (b) square of the linear

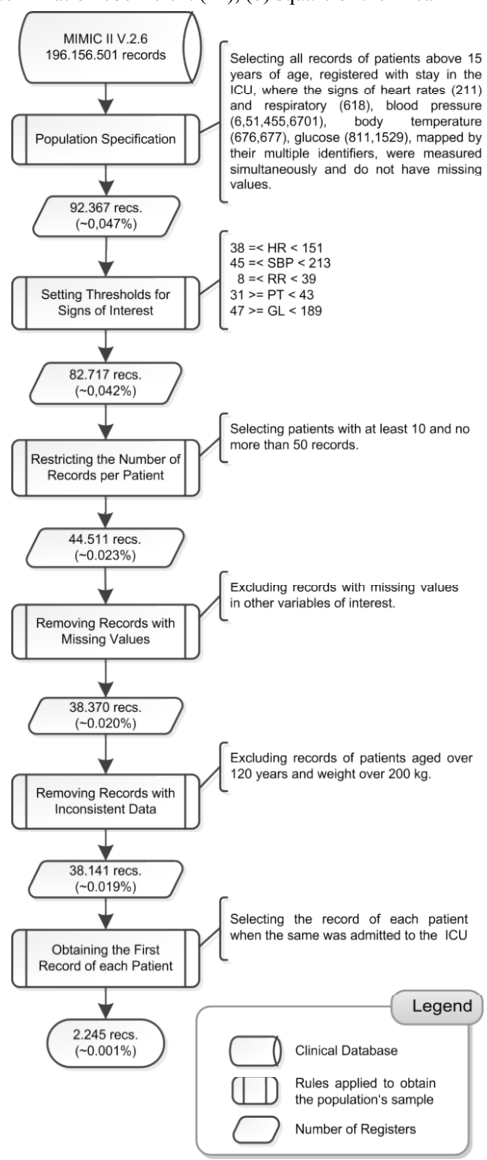


Figure 1 – Process to characterize the population of interest

correlation coefficient (R^{2*}) between response variable and adjusted values. Both measures were used to indicate the data variability explained by the regression model. Table 2 presents the results for each model, with confidence level of 95%.

Table 2 – Metrics to Evaluation of the Regression Models

Regression Model	R^2	R^{2*}
GLM_HR	0.310	0.332
GLM_SBP	0.477	0.511
GLM_RR	0.414	0.489
GLM_PT	0.480	0.480

R^2 = Determination Coefficient, $R^{2*} = COR(y, \hat{\mu})$

As human beings have many variables to be observed, and different individuals react differently to the same drug [13], we believe the results are satisfactory.

It is noteworthy that to predict the respiratory rate we used a *Normal Inverse Model* in the form of Equation (1) with canonical link function defined in Equation (2):

$$\hat{\mu} = \hat{\eta}^{-\frac{1}{2}} \quad (1)$$

where μ is the mean frequency of respiratory rate that we want to model and

$$\hat{\eta} = \hat{\beta}_0 + \sum_{i=1}^{4} \hat{\beta}_i X_i + \sum_{j=2}^{4} \hat{\beta}_{5j} X_{5j} + \sum_{i=6}^{7} \hat{\beta}_k X_k . \quad (2)$$

is the systematic component where β_0 corresponds to the intercept and $\beta_{1\text{-}7}$ to the coefficients of the variables *hr*, *sbp*, *pt*, *gl* and *group*, as well as the interactions between *hr*·*sbp* and *hr*·*gl*, respectively. The *group* variable is used according to the patient classification, given the values of the other predictor variables. Therefore, the X_{5j} variable assumes the value *1* according to the *j* value that specifies the patient group, and the value *0* for all other possible *j* values.

The usage of the *Normal Inverse Model* means that the higher the value of the systematic component, the lower the value of the average estimated respiration rate. Hence, higher values of variables with positive coefficients lead to lower values of *rr*. On the other hand, higher values of variables with negative coefficients, lead to higher values of *rr*.

For the remaining regression models, we used a *Gamma Linear Regression Model* given by Equation (3), whose variance function is more restrained than the *Normal Inverse Model*:

$$\hat{\mu} = \hat{\eta}^{-1} \quad (3)$$

Where μ is the mean frequency of vital signs that we want to model; in this case *hr*, *sbp* and *pt*. Since the workflow to create regression models for all vital signs is similar, these regression models were omitted.

Computational Modeling

To design the baseline patient model we used the AOD paradigm, i.e., a design methodology based on components called *actors* [14]. This methodology represents a formal model of concurrency in which an *actor* is a computational agent that has an independent thread of control and communicates through asynchronous message exchange. We used the *Ptolemy II* modeling tool, which is an extensible AOD-based software framework that supports experimentation, to build the models. Its emphasis is in concurrent components, using well-defined computation models that govern the interaction among these components [15].

The baseline patient model considers the four main vital signs: heart rate (*hr*), respiratory rate (*rr*), arterial blood pressure (*bp*) and peripheral body temperature (*pt*). This patient model consists of the integration among the regression models that represent each one of these vital signs. Such regression models in the AOD paradigm may be modeled as shown in Figure 2, whose regression model is for respiratory rate (*GLM_RR*).

We integrated the regression models to the baseline patient model (see Figure 3) to allow interaction among them and provide the behavioral dynamics to the patient's model. Thus, the user may change the value of a specific vital sign during simulation and automatically the values of other vital signs will be modified according to their respective regression models.

We present part of the elaborated model that incorporates the patient characteristics and the regression models of the four vital signs in Figure 3a. We have highlighted the following key elements: **(1)** the patient model configuration parameters,

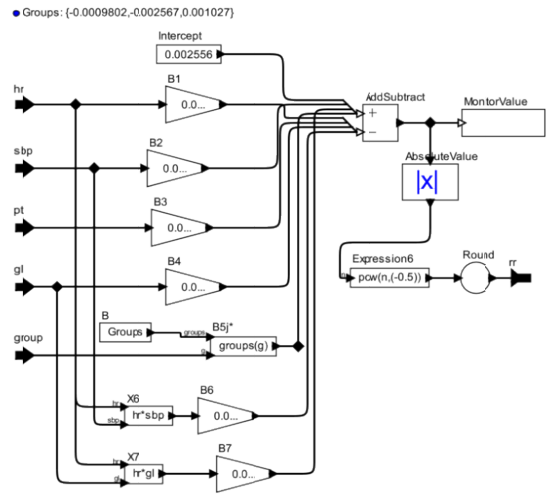


Figure 2 – GLM_RR for the Patient Model

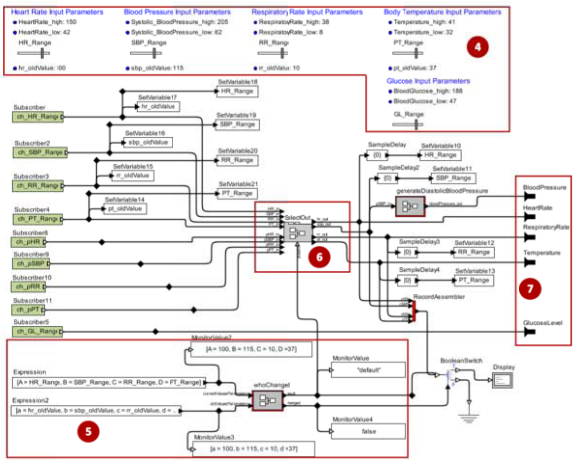
including the *Continuous Director* element that determines their execution semantics; **(2)** the substructure of the model that specifies the vital signs' initial values and physiological parameters represented in the model; **(3)** the elements labeled by the *GLM_<vital_sign>* pattern that denote the regression models for each vital sign modeled. The model parameters are the basis for generating the values for the vital signs provided by the patient model. Moreover, the specifications of the thresholds of each vital sign allow the developer to manipulate them during the simulation, to represent different health conditions for the patient's model. Consequently, the behavior analysis of the MCPS may be carried out for various situations.

We present the second part of the baseline patient model in Figure 3b. The highlighted elements are: **(4)** the parameters that define thresholds for each signal so that these signals remain within the range of values considered in the conception of regression models; **(5)** the logic used to identify which vital sign was changed by the user at a given time during the model simulation; **(6)** the component developed to select the model's correct output according to user intervention; and **(7)** ports that provide the patient model communication interfaces with the medical devices models for data acquisition from these devices.

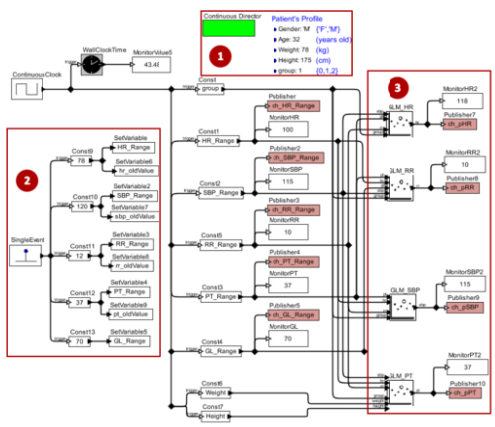
Notice that the patient model only has *output ports* for vital signs and physiological parameters. To receive *feedback control actions* from actuator models, *input ports* must be added to the patient model through the *Ptolemy II* framework. Additionally, the patient model must be adapted to represent the pharmacokinetic model's dynamic behavior corresponding to the drug type to be administered.

Results and Discussion

We used *Diagnostic Plots* to assess the regression models. In Figure 4 we present the envelope (*Normal Q-Q Plot*), leverage points, influential points and residuals versus fitted values plot for *rr*. Due to space constraints, we omit the results from the other vital signs.



(a) Regression models to estimate vital signs data



(b) Submodel to provide information to the medical devices' models

Figure 3 – Baseline Patient Model

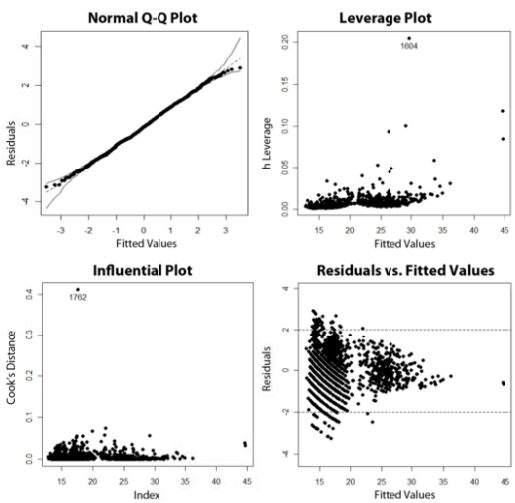


Figure 4 – Diagnostic Plots for the GLM_RR

Envelope plots are useful to check the regression models' fit. Leverage points may interfere in adjusted values close to them and regression coefficients estimate. Influential points may interfere in the model parameters' estimated values. The residuals versus fitted values plot is useful to assess the assumptions of the regression model (e.g., any visible trends would show a dependence of errors on the predictor variable).

After analyzing the diagnostic plots of each regression model, we concluded that the obtained models reasonably explain the vital sign they represent. Thus, the regression models were statistically validated. This means that the synthetic data generated by the patient computational model are compatible with the sample used in statistical analysis. The simulation of the proposed patient model in a specific clinical scenario is out of the scope of this paper.

Conclusion

In this paper, we presented a formal model to provide context information (e.g., profile and vital signs) of a patient, the so-called baseline patient model. We used regression models to generate a set of vital signs (heart rate, respiratory rate, blood pressure and body temperature) that compose the patient model.

The use cases of this patient model are to simulate patients' health conditions to support testing of MCPS. With its extension, it will be possible incorporate it to the context of specific clinical scenarios. This will provide the developer an important tool to identify failures in the system, and will also assist health experts in the strategy planning in the treatment of patients. However, its extension depends on the experience of the MCPS developer, as well as the knowledge of the medical expert that may help in this process.

Our solution can adapt and be used in a variety of clinical scenarios. These scenarios can simulate different medical contexts, helping in the validation process of medical systems. The approach applied to build the baseline patient model may be used to create different patient model types, but will require the use of different samples or clinical data sources. Furthermore, it might be necessary to define other patient use cases to validate the MCPS given that the patient model presented was built using only data of ICU patients.

A potential limitation of this patient model is its restriction to the input values for predictor variables. Whenever these values exceed their thresholds, the accuracy of the regression models used to generate the synthetic data for vital signs might decrease.

In the future, we will show how to extend the patient baseline model for specific clinical scenarios. This will require the incorporation of dynamics of the action of drugs in the patient's body to the model. Thus, the medical device models may act on patients' behavior, causing it to react to such actions. Lastly, we will request that health experts perform external validation of the patient model.

Acknowledgments

The authors would like to thank to CAPES, PPgCC-COPIN and IFPB for the financial support of this research, as well as LANEST-UFCG for the statistical consultancy service.

References

[1] Lee I, Sokolsky O, Chen S, Hatcliff J, et al. Challenges and Research Directions in Medical Cyber-Physical Systems. Proceedings of the IEEE 2012: 100(1): 75-90.

[2] Jiang Z, Pajic M, and Mangharam R. Cyber-Physical Modeling of Implantable Cardiac Medical Devices. Proceedings of the IEEE 2012: 100(1): 122-37.

[3] Agur Z. Interactive clinical trial design. In Faltin F, Kenett R, Ruggeri F, eds. Statistical Methods in Healthcare: John Wiley & Sons, Ltd., 2012; pp. 78-102.

[4] Van Heusden K, Dassau E, Zisser HC, Seborg DE, and Doyle FJ. Control-Relevant Models for Glucose Control Using a Priori Patient Characteristics. IEEE Transactions on Biomedical Engineering, Jul 2012; pp. 1839-1849.

[5] Khan SH, Khan AH, and Khan ZH. Artificial Pancreas Coupled Vital Signs Monitoring for Improved Patient Safety. Arabian Journal for Science and Engineering, Nov 2013; pp. 3093-3102.

[6] Silva LC, Perkusich M, Bublitz FM, Almeida HO, and Perkusich A. A Model-based Architecture for Testing Medical Cyber-physical Systems. In Proceedings of the 29th Annual ACM Symposium on Applied Computing 2014; New York: NY, USA: ACM; pp. 25-30.

[7] Buxton AE, Calkins H, Callans DJ, DiMarco JP, others.

[8] Chobanian AV, Bakris GL, Black HR, Cushman WC, et al. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003: 42(6): 1206-52.

[9] McGee SR. Evidence-based physical diagnosis. In. St. Louis: Saunders Elsevier, 2007; pp. 187-202.

[10] PHYSIONET. Physionet.org. [Online].; 2012 [cited 2012 Nov. 05. Available from: http://physionet.org/mimic2.

[11] Paula GA. Modelos de Regressão com Apoio Computacional São Paulo: IME-USP, 2010.

[12] Gavish. B, Ben-Dov IZ, and Bursztyn M. Linear relationship between systolic and diastolic blood pressure monitored over 24 h: assessment and correlates. J Hypertens 2008: 26(2): 199-209.

[13] Katzung BG, Masters SB, Trevor AJ, eds. Basic & Clinical Pharmacology. 12th ed.: McGraw-Hill, 2012.

[14] Agha G. Actors: a model of concurrent computation in distributed systems Cambridge: MA, USA: MIT Press, 1986.

[15] Berkeley UC. The Ptolemy Project: heterogeneous, modeling and design. 2012.

Address for correspondence

Lenardo Chave e Silva, Embedded Systems and Pervasive Computing Lab., UFCG, 882 Aprígio Veloso Street, Universitário, Campina Grande, PB, Brazil. Email: lenardo.silva@embedded.ufcg.edu.br.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-554

Leveraging Electronic Tablets and a Readily Available Data Capture Platform to Assess Chronic Pain in Children: The PROBE system

Vibha Anand[a,b], Steven J. Spalding[a]

[a] *Center and Department of Pediatric Rheumatology, Pediatric Institute, Cleveland Clinic*
[b] *Department of Quantitative Health Sciences, Lerner College of Medicine of CWRU, Cleveland, OH, USA*

Abstract

The Patient Risks Outcomes and Barriers Evaluation (PROBE) system is developed using a readily available data capture platform (REDCap) and iPads. PROBE performs complete and consistent assessment of pain at every patient visit in pediatric rheumatology practices of our very large healthcare system. Using evidence based clinical guidelines, it combines the following essential elements for care: 1) screening for behavioral risks for chronic pain such as anxiety, sleep deprivation, or painful conditions affecting a caregiver living in child's home, 2) capturing disease activity related measures and enabling 3) clinical decision support. In this demonstration project we describe PROBE and evaluate it for usability in practice. Using PROBE, we observed significant differences for behavioral risk factors in children with juvenile idiopathic arthritis in those who report chronic pain vs. not.

Keywords:
Electronic Tablets; iPad; REDCap; PROBE; CDSS; JIA; Pain; Children

Introduction

Pain is the most distressing aspect of disease in children and can play a predominant role in their everyday lives. Despite pain's impact on both children and their families, the medical community has only recently begun to focus its resources on the investigation and treatment of pain. The field of pediatric rheumatology is not alone in this delay. In June 2011, the Institute of Medicine (IOM) released a landmark report, Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. This report addresses the US healthcare system for its deficiencies in the assessment and management for pain in the United States[1]. The report calls for progress in the collection and reporting of data on pain. It also emphasizes the need to provide complete and consistent assessment of pain as well as the necessity for expanding opportunities in pain research. Children and families who live with juvenile idiopathic arthritis (JIA) can ardently verify the findings of this IOM report.

Children with JIA who experience pain regularly perceive themselves to be more disabled and are more likely to restrict their activity[2-5]. Their pain related to disease commonly persists into adulthood and can significantly impact their productivity and quality of life[6-8]. However, the relationship between disease activity and pain in children and adolescents with rheumatic disease is known to be inconsistent [3,9] but certain behavioral risk factors have shown to influence the pain experience in children with JIA[10-12]. These include sleep deprivation, anxiety, or painful conditions affecting a caregiver in the child's home. Because early identification of risks may help in improving long term patient outcomes, numerous studies have shown that healthcare providers need to implement complete and consistent assessment of pain in their everyday practice[13-18]. Besides, many evidence based guidelines exist for screening and assessing pediatric chronic pain. However, it remains a formidable challenge for most practices to implement these guidelines in routine care [12,19,20] let alone initiate any intervention following screening. Difficulty in implementing guideline components within the workflow and constraints of the clinical workflow are often cited as potential hurdles. In our previous work in general pediatric practices, we have shown that computerized clinical decision support systems (CDSS) can automate screening in waiting rooms [21,22] and alert the providers based on screening. CDSS can be effective tools for case finding [23] and for improving physicians' use of and adherence to care guidelines[23-27]. More recently, we have also shown that electronic tablets are desirable mediums for implementing screening in the waiting rooms, and using these devices, CDSS can gather more complete patient data for making care decisions[28].

Therefore, to address complete and consistent assessments of pain in children and adolescents with JIA at every visit, we have developed an electronic data capture and clinical decision tool - *Patient Risks, Outcomes and Barriers Evaluation (PROBE)*. PROBE is currently being used in three pediatric rheumatology practices of our large health care system. It screens patient families for pain and associated behavioral risk factors in the waiting room using an electronic device (iPad) that communicates with the PROBE back end server in real-time. The objective of this pilot study is to demonstrate - 1) the feasibility of using a readily available data capture platform and electronic tablets for complete and consistent assessment of pain during routine care in pediatric rheumatology practices; 2) that differences are observed on screening in patient reports of chronic pain; and 3) that PROBE can evolve into a viable Health IT platform for pediatric pain management and patient engagement.

Materials and Methods

The PROBE System

The PROBE system consists of three data capture forms, a central server and electronic tablets such as iPads. Access to the data forms is provided using secure wireless network via a

web browser interface. The data forms capture information from patients or caregivers in the waiting room and clinical data from the clinical team in the exam or the work room. The three PROBE forms are - the Patient Screening Questionnaire (PSQ), the Patient Nursing Form (PNF) and the Provider Work form (PWF). Below we describe a workflow using PROBE. Please refer to Figure 1 for details of the workflow.

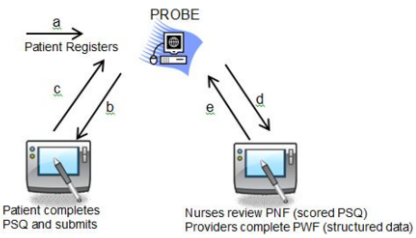


Figure 1 - PROBE Workflow

Workflow: At each patient registration visit (step a) patients or their caregivers are instructed to answer questions on PSQ using an electronic tablet device in the waiting room. Answers to each page of questions are securely recorded in the PROBE central server in real time (b), and no data is ever recorded on the electronic device (c). When responses to screening questions are submitted (d) the nursing form (PNF) is automatically populated with patient answers and scores (e). Inside the exam room, either the provider or the nursing staff completes the PWF (f).

Pain and Behavioral Risk Screeners: Children, adolescents and young adults, ages 3 to 23 years are screened. The following validated screening instruments are implemented using PSQ: *1) Mood disorder screening:* Childhood Depression Index and Screen for Child Anxiety Related Emotional Disorders (SCARED)[29]; *2) Sleep assessment:* Bedtime problems, Excessive daytime sleepiness, Awakening during the night, Regularity and duration of sleep and Snoring (BEARS)[30]; *3) Parental Pain History Questionnaire* (PPHQ)[31]: which is directed towards the caregiver; and *4) Pain Coping Questionnaire:* consisting of *Pain coping efficacy* and *Brief pain inventory scales*[32]. The latter measures weekly pain level on a 0-10 VAS scale, with 0 being the lowest and 10 the highest pain level. Please see Figure 2 below. The *Chronic pain (CP)* report is screened as pain greater than 3 days per week for more than 3 months.

Implementation: PROBE is developed using a readily available electronic data capture platform, the Research Electronic Data Capture (REDCap)[33] platform in collaboration with the Department of Quantitative Health Sciences (QHS) at Lerner Research Institute, Cleveland Clinic. REDCap is a secure data capture tool for research purposes and is supported in our environment. For the PROBE implementation, a REDCap project was created and an anonymous surveys consisting of validated screening questionnaires (*SCARED, BEARS, Brief Pain Inventory Scale, Parental Pain History and Pain Coping and Efficacy* as described above) were implemented for PSQ. The patients do not need login credentials but the front-desk staff enters patients' identifying information into PSQ before handing out the electronic tablet device. At each visit, the front desk staff confirms that the patient is an established JIA patient (from registration) and records the patient's medical record number, age and gender (of the child) and selects the provider and visit type (i.e. for infusion or provider) from a menu item in PSQ. No patient information for race/ethnicity or insurance category is currently recorded in PSQ. Additionally, if the child is between 3 and 12 years of age, the front desk asks the parent or the caregiver to complete the PSQ. Adolescents and young adults (12 – 23 years) complete their own PSQ. The screeners implemented in PSQ have both a parent and a child version, and based on the age entered by the front-desk staff, the appropriate one is automatically presented in PSQ. We use the branching logic capabilities of the REDCap software to implement age-based questions. Using REDCap database functionality, PROBE also automatically assigns a unique system identifier for all three data capture forms and together they constitute a record for the patient's visit. PROBE was initially implemented for a pilot test in one pediatric rheumatology clinic at our main hospital in March 2014. We currently have three practice sites, including two community sites using PROBE. Each site has at least two electronic tablet devices for patient use and they all communicate with the PROBE central server in real-time.

Clinical decision aid tool: PROBE implements clinical decision support via scoring screening questionnaires. When the PSQ form (Figure 2) is submitted by the patient, numeric scores (e.g. anxiety screening) are computed in real-time by the system and displayed to the clinical staff on the nursing form (PNF). Please see Figure 3 for PNF. The providers can access patients' PNF to view scores or any screening information via a secure login to the REDCap website either using an exam room computer or an electronic tablet device. At the conclusion of a visit, providers or their staff document current labs, medications, and disease activity scores. Please see Figure 4 for PWF documentation. More structured data item additions are planned for PWF in the near future. These include orders (labs, medications, imaging) and results from other investigations (e.g. eye exam) to support clinical decision-making. The following types of documentation notes are also planned as additions in the near future for PWF - psycho-social, and sleep evaluation, nutrition, physiotherapy and occupational therapy referrals**.**

At each visit, data from all the three forms are available to the providers and their nursing staff for any further action. However, currently no reminders or prompts to providers or their clinical staff are provided using PROBE. These are planned as future additions as well. Although information for orders, labs and referrals may exist in our electronic medical record (Epic Systems), they may mostly exist as free-text notes, so PROBE offers the capture of structured data items and the ability to use these data for follow up visits using a computer-based clinical decision support system that is also planned for development in the near future.

Analyses: As part of this study, we report descriptive statistics, i.e. number of visits since deployment and patient visit characteristics. We also condcuted bi-variate analyses for an outcome of the self-report of chronic pain with a self-report of anxiety symptoms, sleep impairment issues and parental history of pain as explanatory variables. For purposes of the statistical analyses, we dichotomized (yes or no) the explanatory variables according to the screener scoring thresholds (e.g. anxiety) and questions that were answered in affirmative (e.g. sleep and parental history).

Statistical analyses were performed using Stata Software version 12 (StataCorp, College Station, TX). The study was approved by the Cleveland Clinic Institutional Review Board.

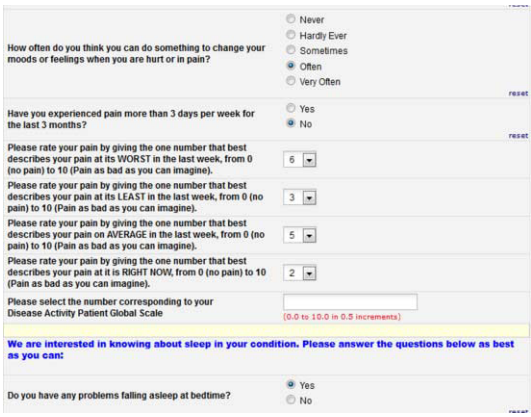


Figure 2 - Patient Screening Questionnaire (PSQ)

Figure 3 - Patient Nursing Form (PNF)

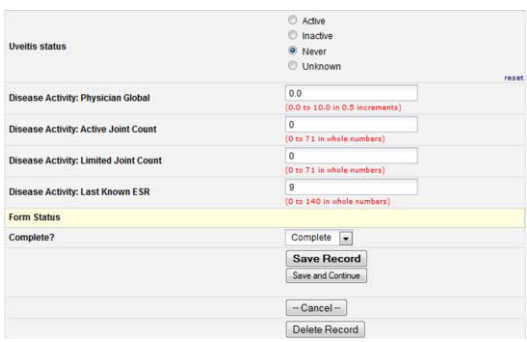


Figure 4 - Provider Work Form (PWF)

Results

Initial results: Over a three-month initial period, PROBE recorded 21 patient visits in the main hospital, our first pilot site. From these visits, we found that patients reporting chronic pain are more likely to report higher average pain scores (>=5) in the last week. They show symptoms of anxiety on as many as 17 of the 41 items screened using three point SCARED scale and are more likely to report sleep issues, e.g. problems falling asleep at bedtime, or feeling sleepy a lot during the day, in school or while driving when compared to those who do not report chronic pain. The patients reporting chronic pain are young adolescents (> 12 years) and more likely to be females with a parental history of pain.

Descriptive Results: As we observed these differences in one practice during routine care, we extended the use of PROBE to the other two practices at our community sites. Subsequent results presented here are from all three practice sites. At the time of this writing, there are 192 patient visits recorded in PROBE. Please see Table 1 below for patient visit characteristics. Please note that in Table 1 there may be missing data for explanatory variables (e.g. sleep) as the patient or the caregiver need not answer all questions at every visit. Similarly, as the system was being deployed and the front-desk staff was getting acquainted with the procedures for recording patient identifying information, there may be some missing values for socio-demographic variables (age, sex). A majority of visits are for female patients (77%) and over 12 years of age (~70%). Patients reported chronic pain in almost half of all visits (49%).

Table 1 – Patient Visit Characteristics

Sex (n=191)	
Female	148 (77.5%)
Age(n=183)	
3 to 7 years	17 (9.3%)
8 to 11 years	39 (21.3%)
12 to 18 years	107 (58.5)
18 to 23 years	18 (9.6%)
Chronic Pain Report (n=181)	
No	92 (50.8%)
Yes	89 (49.2%)
Anxiety Report (n=192) §	
No	161 (83.9%)
Yes (score >= 25 and < 30)	11 (5.7%)
Yes (score >= 30)	20 (1%)
Sleep Issues Report (n=171)¥	
No	59 (34.5%)
Yes	112 (65.5%)
Parental History of Pain Report (n=154)£	
No	25 (16.2%)
Yes	129 (83.8%)

scoring or interpretation according to § SCARED; ¥ BEARS; £ PPHQ

Bi-variate Analyses: Using the chi-square test, the report of one or more sleep impairment issues was significantly associated with the report of chronic pain ($X^2 = 15.7$, p-value < 0.0001). The total score on anxiety symptoms were not significantly associated with the report of chronic pain, as was the parental history of pain report from this pilot study data.

However, we observed differences based on specific anxiety symptoms reported on the SCARED screening tool. In the bivariate analysis, the presence of panic disorder or significant somatic symptoms (score of 7 or more), separation anxiety (score of 5 or more), and significant school avoidance (score of 3 or more) on specific questions are significantly associated with the patient report of chronic pain (X2 = 4.24, p-value = 0.039) and (X2 = 7.11, p-value = 0.008) respectively.

Similarly, we observed differences in responses to pain coping questions for patients reporting chronic pain versus no chronic pain. In general, those reporting chronic pain reported higher pain score levels (4 to 5 on a 10 point VAS) as opposed to 2 to 3 on a 10 point VAS and lower coping levels (e.g. "When hurt or in pain for a few hours or a few days how often do you think you can do something to change it?" The responses were never, or hardly ever for patients reporting chronic pain versus often or very often for those reporting no chronic pain).

Discussion

We have demonstrated that by using a readily available data capture platform and by leveraging electronic tablets such as iPads as an interface, we can create a tool for complete and consistent assessment of pain at every visit in children and adolescents. Furthermore, our analyses show that there are differences in behavioral risks for patients who report chronic pain versus those who do not and we have demostrated this in a real-world clinical setting.

In this pilot work, we have also shown that the PROBE system can effectively screen patients or their caregivers in pediatric rheumatology waiting rooms. Our results show that impaired sleep is a strong predictor for a chronic pain report in children and adolescents with JIA. However, as we write this, we are collecting more data to develop and evaluate a comprehensive prediction model for chronic pain in this population. Simultaneosuly, we are working on developing PROBE into an automated computer-based clinical decision support system. We hope to evaluate it in near future for patient and system based interventions for pain management in clinical studies.

Because the use of PROBE in routine care can help identify patients' risk factors, i.e. chronic pain vs. no chronic pain group patients, we hypothesize that it can be used to support patient engagement activities, for example to ask more detailed questions about a patient's painful condition, their coping style and/or to provide educational content. Potentially, these activities can promote patient-clinician interaction and lead to informed clinical decisions such as a referral for sleep medicine, psycho-social evaluation or cognitive behavioral therapy (for example for pain coping). Therefore, PROBE based clinical decision support can be used to manage patients who are at higher risk of chronic pain prospectively, however all this needs to be studied in well-designed clinical trials.

As with all such work, there are limitations to ours too. First, the PROBE system is limited by use of REDCap and its built-in programming functionality (e.g. unable to query data on a survey form in real-time). In that respect, the current system is a prototype or demonstration project. Secondly, because of the above limitations, we have not built automated clinical decision support (e.g. prompts and reminders) functionality into the current system. Therefore, our future work would involve migrating PROBE to a more robust data capture and query platform so that we can build robust patient engagement tools, and provide clinical decision support. Lastly, our bi-variate analyses of pilot data may suffer from sample and confounding bias. Regardless, our demonstration project highlights the importance of using eHealth enabled tools in pediatric specialty care practices.

Acknowledgments

The authors would like to thank the front-desk and nursing staff at the Pediatric Rheumatology clinics at Cleveland Clinic. A special thanks to Carmen Howard for assistance with the initial pilot. The authors would also like to thank QHS, especially Heather Kren for assistance with programming and supporting the PROBE system in REDCap.

References

[1] Committee on Advancing Pain Research C, Education, Institute of M. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research.* Washington, D.C.: The National Academies Press; 2011.

[2] Sawyer MG, Carbone JA, Whitham JN, et al. The relationship between health-related quality of life, pain, and coping strategies in juvenile arthritis--a one year prospective study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2005;14(6):1585-1598.

[3] Schanberg LE, Gil KM, Anthony KK, Yow E, Rochon J. Pain, stiffness, and fatigue in juvenile polyarticular arthritis: contemporaneous stressful events and mood as predictors. *Arthritis and Rheumatism.* 2005;52(4):1196-1204.

[4] Shaw KL, Southwood TR, Duffy CM, McDonagh JE. Health-related quality of life in adolescents with juvenile idiopathic arthritis. *Arthritis and Rheumatism.* 2006;55(2):199-207.

[5] Limenis E, Grosbein HA, Feldman BM. The relationship between physical activity levels and pain in children with juvenile idiopathic arthritis. *The Journal of rheumatology.* 2014;41(2):345-351.

[6] Eyckmans L, Hilderson D, Westhovens R, Wouters C, Moons P. What does it mean to grow up with juvenile idiopathic arthritis? A qualitative study on the perspectives of patients. *Clinical rheumatology.* 2011;30(4):459-465.

[7] Shaw KL, Southwood TR, McDonagh JE, British Paediatric Rheumatology G. Transitional care for adolescents with juvenile idiopathic arthritis: a Delphi study. *Rheumatology (Oxford, England).* 2004;43(8):1000-1006.

[8] Thastum M, Zachariae R, Herlin T. Pain experience and pain coping strategies in children with juvenile idiopathic arthritis. *The Journal of rheumatology.* 2001;28(5):1091-1098.

[9] April KT, Feldman DE, Platt RW, Duffy CM. Comparison between Children with Juvenile Idiopathic Arthritis (JIA) and their parents concerning perceived Quality of Life. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2006;15(4):655-661.

[10] 10. Palermo TM, Eccleston C, Lewandowski AS, Williams AC, Morley S. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. *Pain.* 2010;148(3):387-397.

[11] Palermo TM, Kiska R. Subjective sleep disturbances in adolescents with chronic pain: relationship to daily functioning and quality of life. *The journal of pain : official journal of the American Pain Society.* 2005;6(3):201-207.

[12] Schanberg LE, Anthony KK, Gil KM, Lefebvre JC, Kredich DW, Macharoni LM. Family pain history predicts child health status in children with chronic rheumatic disease. *Pediatrics.* 2001;108(3):E47.

[13] Reigada LC, Hoogendoorn CJ, Walsh LC, et al. Cross-Sectional Analysis of Anxiety Symptoms and Disease Severity in Children and Adolescents with Crohn's Disease. *Journal of pediatric gastroenterology and nutrition.* 2014.

[14] Seid M, Huang B, Niehaus S, Brunner HI, Lovell DJ. Determinants of health-related quality of life in children newly diagnosed with juvenile idiopathic arthritis. *Arthritis care & research.* 2014;66(2):263-269.

[15] Armbrust W, Kaak JG, Bouma J, et al. Assessment of disease activity by patients with juvenile idiopathic arthritis and the parents compared to the assessment by pediatric rheumatologists. *Pediatric rheumatology online journal.* 2013;11(1):48.

[16] Stinson JN, Jibb LA, Lalloo C, et al. Comparison of Average Weekly Pain Using Recalled Paper and Momentary Assessment Electronic Diary Reports in Children with Arthritis. *The Clinical journal of pain.* 2014.

[17] Millman P, Pilossyan L, Kimura Y, Weiss JE, Nyirenda T, Janow GL. A34: seeing the forest through the trees: predictors of inflammatory causes of joint pain in new patients. *Arthritis Rheumatol.* 2014;66 Suppl 11:S53.

[18] Guzman J, Gomez-Ramirez O, Jurencak R, et al. What Matters Most for Patients, Parents, and Clinicians in the Course of Juvenile Idiopathic Arthritis? A Qualitative Study. *The Journal of rheumatology.* 2014.

[19] Cramton RE, Gruchala NE. Managing procedural pain in pediatric patients. *Current opinion in pediatrics.* 2012;24(4):530-538.

[20] O'Donnell FT, Rosen KR. Pediatric pain management: a review. *Missouri medicine.* 2014;111(3):231-237.

[21] Anand V, Biondich PG, Liu G, Rosenman M, Downs SM. Child Health Improvement through Computer Automation: the CHICA system. *Stud Health Technol Inform.* 2004;107(Pt 1):187-191.

[22] Anand V, Carroll AE, Downs SM. Automated primary care screening in pediatric waiting rooms. *Pediatrics.* 2012;129(5):e1275-1281.

[23] Carroll AE, Biondich P, Anand V, Dugan TM, Downs SM. A randomized controlled trial of screening for maternal depression with a clinical decision support system. *Journal of the American Medical Informatics Association : JAMIA.* 2012.

[24] Carroll AE, Anand V, Dugan TM, Sheley ME, Xu SZ, Downs SM. Increased Physician Diagnosis of Asthma with the CHICA Decision Support System. *Pediatr Allergy Immunol Pulmonol.* 2012;August 2012.

[25] Carroll AE, Bauer NS, Dugan TM, Anand V, Saha C, Downs SM. Use of a computerized decision aid for ADHD diagnosis: a randomized controlled trial. *Pediatrics.* 2013;132(3):e623-629.

[26] Carroll AE, Biondich PG, Anand V, et al. Targeted screening for pediatric conditions with the CHICA system. *Journal of the American Medical Informatics Association : JAMIA.* 2011;18(4):485-490.

[27] Carroll AE, Bauer NS, Dugan TM, Anand V, Saha C, Downs SM. Use of a computerized decision aid for developmental screening and surveillance: A randomized controlled trial. *JAMA Pediatrics.* 2014.

[28] Anand V, McKee SJ, Dugan TM, Downs SM. Leveraging Electronic Tablets for General Pediatric Care: A Pilot Study. *Appl Clin Inform.* 2015;6(1):1-15.

Address for correspondence

Vibha Anand, PhD, Assistant Professor of Pediatrics,
Department & Center for Pediatric Rheumatology and Immunology, Pediatric Institute, Cleveland Clinic, Department of Quantitative Health Sciences, Lerner College of Medicine of Case Western Reserve University, 9500 Euclid Ave JJN3-01Cleveland, Ohio 44195, USA

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-559

Normalization of Phenotypic Data from a Clinical Data Warehouse: Case Study of Heterogeneous Blood Type Data with Surprising Results

James J. Cimino

Laboratory for Informatics Development, the Clinical Center of the US National Institutes of Health, Bethesda, Maryland, USA

Abstract

Clinical data warehouses often contain analogous data from disparate sources, resulting in heterogeneous formats and semantics. We have developed an approach that attempts to represent such phenotypic data in its most atomic form to facilitate aggregation. We illustrate this approach with human blood antigen typing (ABO-Rh) data drawn from the National Institutes of Health's Biomedical Translational Research Information System (BTRIS). In applying the method to actual patient data, we discovered a 2% incidence of changed blood types. We believe our approach can be applied to any institution's data to obtain comparable patient phenotypes. The actual discrepant blood type data will form the basis for a future study of the reasons for blood typing variation.

Keywords: Clinical Data Repositories, Phenotype Detections, Blood Typing

Introduction

Clinical data warehouses are becoming a common tool for providing access to electronic health record data for various forms of reuse.[1] However, the reuse of historical data can be problematic, especially when data are pooled from multiple sources for which minimal documentation and metadata exist.[2] A challenge for clinical research informatics is to provide users with ways to reconcile such heterogeneous data.[2] For example, if one source codes gender as "male or female" and another source codes gender as "male, female, or other", summarizing gender data across sources becomes problematic.

While normalzation procedures are standard for genetic sequence data,[3] they have been less well defined for phenotypic data, typically involving ad hoc mappings between patient characteristics.[4] The purpose of this paper is to present a method for normalization of heterogeneous data by reducing them to specific, well-characterized phenotypic traits. We take as our example common ABO and Rh blood typing such as is typically tested in blood bank laboratories.

Background

The Biomedical Translational Research Information System

The Clinical Center of the US National Institutes of Health (NIH) is a hospital in Bethesda, Maryland that has served as a site for clinical research since 1953. Over the years, data from clinical studies at NIH have been captured in a variety of clinical trials data management systems, as well as two electronic health records – one in operation from 1976 to 2004 and one in operation since 2004. These data, on over 500,000 human subjects from over 50 source systems are collected into a single database that forms the core of the NIH's Biomedical Translational Research Information System (BTRIS).[5] All terms from source systems are assigned unique concept identifiers in BTRIS's Research Entities Dictionary (RED), which is a unified ontology that includes hierarchical classifications of similar terms (such as tests that measure the same substance or medications that contain the same ingredients). BTRIS users select specific terms or classes of terms from the RED to use in retrieving identified data on their own clinical studies, as well as de-identified data across all clinical studies.[5]

Case Study : ABO and Rh Blood Type Antigens

Human red blood cells express a wide variety of antigens. Three in particular (A, B and Rh) are commonly identified in clinical laboratories for purposes such as cross-matching blood for transfusion. Blood donors and blood recipients are characterized as having A, B, AB or O blood and as being Rh positive or negative. Except in rare instances, an individual's blood type does not change over his or her lifetime.

Today, most blood banks report these antigens together as the result of a single test; e.g., A+, B-, AB+, O-, etc. However, in the past, these results were reported across two tests: the ABO test to report types A, B, AB and O, and the Rh test with the possible results "positive" and "negative". Even earlier, the ABO test results were reported as two separate tests for the A and B antigens. A patient could have a positive results for both of these tests (type AB), one of these tests (type A or B), or neither of these tests (implying type O).

This heterogeneity has been well-documented as a challenge for dealing with pooling data from, or sharing them between, multiple health care sites.([6], also: Huff SM, personal communication; inspiring [7]) As a repository of data from multiple sites over 40 years, BTRIS often presents users with this type of challenge. We chose it as a case study for the normalization of heterogeneous data, since proper normalization should result in patients having consistent blood types over time, in effect serving as their own gold standards.

Methods

Approach to Phenotype Normalization

Our approach to phenotypic characterization involves reducing complex findings to their most atomic forms and them assembling them in canonical ways into more complex phenotypic patterns. In the case of blood typing, we consider that each test provides evidence for the presence or absence of at

least one red blood cell antigen. This can range from a single antigen (A, B, or Rh) to all of them (for example, type "AB+" indicates the presence of all three and type "O-" indicates the absence of all three).

Preliminary Analysis of Result Types

The first step in our normalization process was to identify the relevant tests panels, individual tests and actual results reported by those tests. We used the BTRIS Limited Data Set function[8] to retrieve de-identified data from BTRIS, using appropriate terms selected from the RED.

The second step was to review each unique panel-test-result triple to determine the antigenic evidence it provides. Each result was tagged with the letters A, B and R, with presence indicated by an upper-case letter and absence indicated by a lower case letter. Thus, the a B-Antigen test with the result "Positive" was labeled as "B" and a Type and Crossmatch test with the result "O+" was labeled as "abR". A MUMPS data structure (PC-MUMPS, DataTree Inc., Waltham, MA), was created for each result and its assigned antigens.

Analysis of Patient Data for Phenotype Consistency

In the third step, the relevant data of individual patients were characterized based on the assigned antigens. Pooled evidence for the presence or absence of each antigen was stored for each patient in a second MUMPS global, such that a patient would typically have three letters (one each of "A" or "a", "B" or "b", and "R" or "r"). In the fourth step, we reviewed the results for each patient to determine situations where an individual did not have exactly one of each letter.

All data were obtained with oversight of the NIH Office of Human Subjects Research Protection (Agreement Number BTRIS_2014_835_CIMINO_J_CC). Only those data that did not require permission for reuse from the original investigators, as per NIH policy, were retrieved. The patients' birthdates and test dates were included in the data but all other potentially identifying information was removed, as per NIH policy for limited-use data sets.

Results

Preliminary Analysis of Result Types

A search of the RED for tests with names containing the phrase "blood group" identified 644 terms, including three appropriate term classes as shown in Figure 1: *ABO Grouping and/or Rh Antigen Phenotyping Intravascular Test*, *Blood Group Antigen Blood Typing Blood Bank Test*, and *RH Blood Group System Antigen (Rh Factor) Test*. When BTRIS was queried with these three terms, 593,637 test results were found on 43,485 patients (see Figure 2). The data included results from 139 tests in 66 test panels, with 334 unique panel-test pairs that reported 3946 unique results (Figure 3).

Review of all panels and tests identified 21 panels and 59 tests within those panels that provide information on ABO and Rh blood typing. The data set included 1452 unique test results for these tests (many of which were misspellings, as in [6]). Each was reviewed manually to assign antigenic evidence. Table 1 shows a sample of panels, tests, results and assigned antigenic evidence.

Assignment of Phenotype Based on Antigenic Evidence

After removal of irrelevant panels and tests, the patient data included 165,981 panel events with 307,884 test results for 43,485 patients. Pooling of all antigenic evidence for each

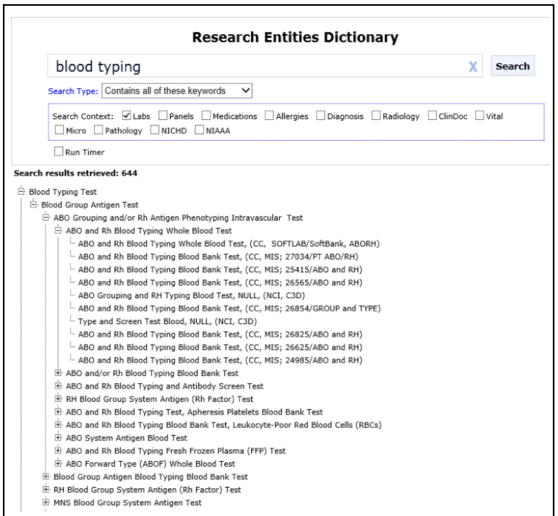


Figure 1: Blood type terms in the Research Entities Dictionary. A search for the phrase "blood typing", restricted to the laboratory domain identified one class ("Blood Typing Test") that subsumes 643 additional terms. The hierarchy is partially expanded.

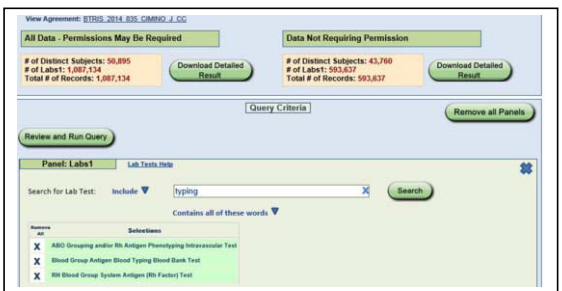


Figure 2: Query screen of BTRIS Limited Data Set application, showing three search terms at bottom and query summaries at top. Only data from the right ("Data Not Requiring Permissions") were downloaded.

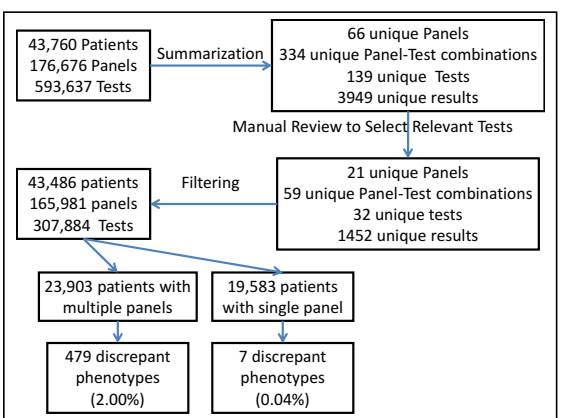


Figure 3: Summary of analysis of initial data set.

Table 1: Panel/test/result combinations with interpretations

Panel	Tests	Result	Interpretation
ABO GRP-RH TYPE	ABO GRP-RH TYPE	O POSITIVE	abR
ABO GRP-RH TYPE	ABO GRP-RH TYPE	0 POS	abR
ABO GRP-RH TYPE	ABO GRP-RH TYPE	A POSITIVE	AbR
ABO GRP-RH TYPE	ABO GRP-RH TYPE	A NEG	Abr
ABO GRP-RH TYPE	ABO GRP-RH TYPE	A NEGATIVE	Abr
ABO GRP-RH TYPE	ABO GRP-RH TYPE	AB NEG	ABr
ABO GRP-RH TYPE	ABO GRP-RH TYPE	B POS	aBR
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	0	a
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	1+	A
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	2+	A
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	3+	A
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	4+	A
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - A	M4	A
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - B	0	b
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - B	4+	B
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - Rh	NEG	r
ABO Group and Rh Type [ABORH]	ABO Group and Rh Type [ABORH] - Rh	POS	R

Panels are collections of one or more tests, each of which is associated with a result. We interpret the results to correspond to specific phenotypic characteristics: A=Antigen present; B=B antigen present, AB=A and B antigens present; a=A antigen absent; b=B antigen absent; ab=Neither A nor B antigens present ("Type O"); R=Rh antigen present ("Rh Positive"), r=Rh antigen absent ("Rh Negative"),

patient identified 32 different phenotypes, of which 8 were complete (ABO and Rh designations), 6 were incomplete (ABO or R designation only) and 18 were discrepant (multiple ABO and/or Rh designations). Table 2 shows the antigenic evidence, and counts for each phenotypic designation.

Review of Aberrant Phenotypes

In all, 531 of the 43,485 patients (1.22%) had aberrant phenotypes, based on discrepant laboratory results. Some examples of their test results are shown in Table 3. Given that random, patient-independent laboratory errors could account for some of these discrepancies, and that the frequency of errors would be proportional to the number of tests run, we examined the frequency aberrant phenotypes versus number of test panels performed (Figure 4). The Pearson coefficient for this association is 0.7127 (P<.00001); however, the slope of the regression line is only 0.04.

Initially, we did not consider patients with a single test panel, since we were looking for discrepancies between panels. Doing so would result in discrepancies in 479 of 23,903 patients (2.00%). However, we were surprised to find that seven of 19,583 patients (0.004%) with a single test panel had discrepancies, so although rare, this situation was not impossible. This led us to examine intra-panel discrepancies. We found that 109 patients (including the seven mentioned above) had 588 intra-panel discrepancies (see Table 3).

Discussion

Informatics Implications

As previously noted, heterogeneity of data syntax and semantics in clinical repositories complicates analysis of data aggregated from multiple sources over long time spans. The approach described here recasts clinical findings in terms of atomic phenotypic characteristics that are institution- and laboratory-independent. The case here concerns the presence or absence of specific human red blood cell antigens, but we believe this approach can be applied to many other data types, such as microorganism antigens and antibodies and clinical findings that are sometimes reported as collections of postcoordinated terms and sometimes as a single, precoordinated term. This approach has potential for use with other repositories or networks of repositories with heterogeneous data.

Biomedical Implications

We were surprised at the incidence of intra-patient blood type variation among patients with multiple sets of test results (2%). Human blood antigens do not normally change during a patient's lifetime, except in unusual circumstances such as transplantation,[9] leukemia[10, 11] and, very rarely, viral infection[12]. It is possible – in fact probable – that some of this variation was due to laboratory error. The correlation (r=.72) between numbers of tests performed and likelihood is consistent with an independent influence such as systematic error; however the slope of the correlation (0.04) indicates that this would only account for about 4% of the variance. The data collected in this study are insufficient for determin-

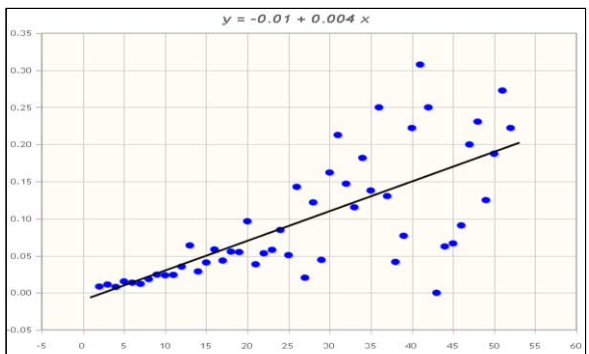


Figure 4: Pearson's correlation between number of tests performed and incidence of at least one discrepant result between the tests. (http://www.mathportal. org/ calculators /statistics-calculator/correlation-and-regression-calculator.php)

Table 2: Antigen patterns found in test results, classified as Complete (corresponding to an actual real phenotype found in nature), Incomplete (lacking sufficient information to determine the phenotype) or Aberrant (not corresponding to an actual phenotype)

Antigenic Evidence	Phenotype	# Patients
Complete		
abR	O+	17132
AbR	A+	13925
aBR	B+	4710
abr	O-	2538
Abr	A-	2316
ABR	AB+	1441
aBr	B-	645
ABr	AB-	214
Incomplete		
r	-	10
R	+	8
ab	O	7
Ab	A	5
AB	AB	1
aB	B	1
bR	+	1
Aberrant		
AabR	(aberrant)	132
abRr	(aberrant)	89
AbRr	(aberrant)	67
aBbR	(aberrant)	51
AaBbR	(aberrant)	50
AabRr	(aberrant)	28
ABbR	(aberrant)	24
aBRr	(aberrant)	19
AaBR	(aberrant)	17
Aabr	(aberrant)	13
aBbr	(aberrant)	11
ABRr	(aberrant)	7
AaBbRr	(aberrant)	6
aBbRr	(aberrant)	6
ABbr	(aberrant)	6
AaBbr	(aberrant)	3
ABbRr	(aberrant)	2

ing what is likely to be a multifactorial cause for the variations. Discovering them will require a future study in which the entire patient records are accessed.

Conclusions

The reduction of heterogeneous laboratory results to their atomic phenotypic characteristics was an effective approach for normalizing patient test results from a longitudinal repository with 40 years of data from disparate sources systems. The normalization makes some patient characteristics explicit (e.g., by inferring Type O from the absence of A and B) as well as highlighting data that are discrepant. The approach seems extensible to other domains and repositories. The incidental finding of intra-patient blood type variation reminds us that the re-use of data constantly holds challenges and surprises, and that every question we seek to resolve seems to produce more new questions than answers.

Acknowledgment

This work was supported by intramural research funds from the NIH Clinical Center and the National Library of Medicine.

References

[1] Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: knowledge discovery in a clinical data warehouse. Proc AMIA Annual Fall Symp. 1997:101-5.

[2] Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ, Saltz JH. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013 Aug;51(8 Suppl 3):S30-7.

[3] Xuan J, Wang Y, Hoffman E, Clarke R. Cross phenotype normalization of microarray data. Front Biosci (Elite Ed). 2010 Jan 1;2:171-86.

[4] Li D, Simon G, Chute CG, Pathak J. Using association rule mining for phenotype extraction from electronic health records. AMIA Jt Summits Transl Sci Proc. 2013 Mar 18;2013:142-6.

[5] Cimino JJ, Ayres EJ, Remennik L, Rath S, Freedman R, Beri A, Chen Y, Huser V. The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): Design, Contents, Functionality and Experience to Date. Journal of Biomedical Informatics, 2014 Dec;52:11-27.

[6] Abhyankar S, Demner-Fushman D. A simple method to extract key maternal data from neonatal clinical notes. Proc AMIA Annual Fall Symp. 2013 Nov 16;2013:2-9.

[7] Baorto DM, Cimino JJ, Parvin CA, Kahn MG. Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC). Inter Journal Med Inform. 1998;51(1):29-37.

[8] Cimino JJ, Ayres EJ, Beri A, Freedman R, Oberholtzer E, Rath S. Developing a self-service query interface for reusing de-identified electronic health record data. Stud Health Technol Inform. 2013;192:632-6.

[9] Bronson WR. McGinniss MH, Morse EE. Hematopoietic graft detected by a change in ABO group. Blood. 1964 Feb;23:239-49.

[10] van der Hart, van der Veer, van Loghem J. Change of blood group B in a case of leukaemia. Vox Sang. 1962 Jul-Aug;7:449-53.

[11] Hocking DR. Blood group change in acute myeloid leukaemia. Med J Aust. 1971 Oct 30;2(18):902-3.

[12] Sherman LA, Silberstein LE, Berkman EM. Altered blood group expression in a patient with congenital rubella infection. Transfusion. 1984 May-Jun;24(3):267-9.

Address for correspondence:

James J. Cimino, M.D.
Director, Informatics Institute
University of Alabama at Birmingham School of Medicine
1720 2nd Ave South
Birmingham, AL 35294-0113 USA
ciminoj@uab.edu

Table 3: Sample results showing discrepant blood type results for the same patient (dates are approximate for privacy reasons). Adjacent rows are results from tests in the same test panel; shaded rows separate panels for the same patient; white rows separate data from different patients. Note that last two subjects show discrepancies within a single panel, resulting in aberrant blood type interpretations.

Subject	Date	Panel	Test	Result	Antigens	Interp.
59	1/31/1989 2:34	TYPE & SCREEN	ABO & RH	O POSITIVE	abR	abR (O+)
59	1/31/1989 8:52	TYPE & SCREEN	ABO & RH	A POSITIVE	AbR	AbR (O-)
724	1/24/1989 9:55	TYPE & SCREEN	ABO & RH	O NEG	abr	abr (O-)
724	2/13/1989 2:22	TYPE & SCREEN	ABO & RH	O POS	abR	abR (O+)
986	1/2/1999 1:25	Type and Antibody Screen	ABO Group and Rh Type - Rh	POS	R	ABR (AB+)
986	1/2/1999 1:25	Type and Antibody Screen	ABO Group and Rh Type - A	4+	A	
986	1/2/1999 1:25	Type and Antibody Screen	ABO Group and Rh Type - B	4+	B	
986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - Rh	POS	R	AbR (A+)
986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - A	4+	A	
986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - B	0	b	
1090	1/2/2002 12:39	Type and Antibody Screen	ABO Group and Rh Type - ABO	A	Ab	AbR (A+)
1090	1/2/2002 12:39	Type and Antibody Screen	ABO Group and Rh Type - Rh	POS	R	
1090	1/2/2002 12:39	Type and Antibody Screen	ABO Group and Rh Type - A	4+	A	
1090	1/2/2002 12:39	Type and Antibody Screen	ABO Group and Rh Type - B	0	b	
1090	1/28/2003 1:29	Type and Antibody Screen	ABO Group and Rh Type - ABO	B	aB	aBR (B+)
1090	1/28/2003 1:29	Type and Antibody Screen	ABO Group and Rh Type - Rh	POS	R	
1090	1/28/2003 1:29	Type and Antibody Screen	ABO Group and Rh Type - A	0	a	
1090	1/28/2003 1:29	Type and Antibody Screen	ABO Group and Rh Type - B	4+	B	
2185	1/11/1991 6:43	TYPE & SCREEN	ABO & RH	O NEG	abr	Rabr (O+/O-)
2185	1/11/1991 6:43	TYPE & SCREEN	ABO & RH	O POS	abR	
3986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - ABO	AB	AB	ABRb (AB+/A+)
3986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - Rh	POS	R	
3986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - A	4+	A	
3986	1/18/2000 1:24	Type and Antibody Screen	ABO Group and Rh Type - B	0	B	

A=Antigen present; B=B antigen present, AB=A and B antigens present; a=A antigen absent; b=B antigen absent; ab=Neither A nor B antigens present ("Type O"); R=Rh antigen present ("Rh Positive"), r=Rh antigen absent ("Rh Negative")

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-564

Automatic Selection of Clinical Trials Based on A Semantic Web Approach

Marc Cuggia [abg], Boris Campillo-Gimenez [acg], Guillaume Bouzille [abg], Paolo Besana [ag], Wassim Jouini [ag], Jean-Charles Dufour [d], Oussama Zekri [c], Isabelle Gibaud [e], Cyril Garde [f], Regis Duvauferier [bg]

[a] INSERM, U1099, Rennes, F-35000, France, [b] CHU Rennes, CIC Inserm 1414, Rennes, F-35000, France
[c] CRLCC Eugène Marquis, [d] LERTIM –Université de la méditerannée, [e]SIB, [f]ENOVACOM, Université de Rennes 1, [g]LTSI, Rennes, F-35000, France

Abstract

Recruitment of patients in clinical trials is nowadays preoccupying, as the inclusion rate is particularly low. The main identified factors are the multiplicity of open clinical trials, the high number and complexity of eligibility criteria, and the additional workload that a systematic search of the clinical trials a patient could be enrolled in for a physician. The principal objective of the ASTEC project is to automate the prescreening phase during multidisciplinary meetings (MDM). This paper presents the evaluation of a computerized recruitment support systems (CRSS) based on semantic web approach. The evaluation of the system was based on data collected retrospectively from a 6 month period of MDM in Urology and on 4 clinical trials of prostate cancer. The classification performance of the ASTEC system had a precision of 21%, recall of 93%, and an error rate equal to 37%. Missing data was the main issue encountered. The system was designed to be both scalable to other clinical domains and usable during MDM process.

Keywords:

Clinical research, Decision support system, Semantic interoperability, automatic reasoning.

Introduction

Clinical trials (CTs) are the gold standard for testing therapies or new diagnosis techniques that would improve clinical care. CTs often rely on adequate sample sizes, but often remain incomplete or are cancelled due to missed recruitment targets within a certain timeframe and financial cost. The recruitment process faces many barriers that have already been well identified in the literature. The automation of the patient screening process by computerized recruitment support systems (CRSS) should address some of the recruitment barriers. In a review paper, we analysed the advantages and drawbacks of each CCRS described in the literature[1]. Based on this analysis, we developed a system based on the 3 following principles:

1. Electronic health records (EHRs) and eligibility criteria (EC) are usually written by humans for humans. Consequently, their formalization for an automatic screening system can prove to be challenging. An CRSS should rather process structured and coded data. Free text formulation of patient data and eligibility criteria results in too many ambiguities. Despite the efficiency of NLP methods, coded and structured data with a formal semantic remain the best situation to successfully develop and apply automatic reasoning methods.
2. To be scalable, a CRSS should be connected to any kind of clinical EHR source. Indeed, data from the same patient might be scattered in different hospital information systems. To address this issue, an CRSS should adopt a service-oriented architecture and use semantic interoperability standards to communicate with the different data sources.
3. Similarly to any decision support system, CRSS should provide useful information about the eligibility status of patients strategically at the right time and place, such as when physicians decide on treatment plans. For instance in oncology, this decision could be taken during MultiDisciplinary Meetings (MDM).

ASTEC (Automatic Selection of clinical Trials based on Eligibility Criteria) is a French national research project that aims to develop an CRSS designed on the 3 principles mentioned above. This system has been tested by the Centre Eugene Marquis (CEM), a Regional Comprehensive Cancer Centre located at Rennes (Brittany,France). In this paper, we present the evaluation of the system.

Background

Multidisciplinary meeting and the recruitment process in clinical trials: MDM are an integrated team approach to health care in which medical and allied healthcare professionals consider all relevant treatment options and develop collaboratively an individual treatment plan for each patient. That is, MDM is about all relevant health professionals discussing options and making joint decisions about treatment and supportive care plans, taking into account the personal preferences of the patient. In France, multidisciplinary decisions are mandatory for all oncology patients.

Decision support systems for recruitment: For over 25 years, many attempts have been made to develop methods and tools supporting the recruitment process. More recently, projects with new technologies recent are arising, such as the Biomedical Translational Research Information System (BTRIS) developed at NIH to consolidate clinical research data [2]. It is intended to simplify data access to and analysis of data from active clinical trials and to facilitate reuse of existing data to answer new questions. The EHR4CR [3] project supports the feasibility, exploration, design and execution of clinical studies and long-term surveillance of patient populations by providing services supported by an european infrastructure connected to hospital Clinical Data Warehouses. TRANSFoRm project has similar objectives but focused on primary care patient data. STRIDE [4] is a platform supporting clinical and translational research consisting of a clinical data warehouse, an application development framework for building research data management applications and a biospecimen data management system. The ObTiMA system relies on OWL and SWRL to perform semantic mediation between heterogeneous data sources [5]. MATE [6] is an interactive computer system to facilitate explicit, evidence-based decision-making in MDM for

breast cancer care. MATE [7] provides prognostication and risk assessment and also flags up patients eligible for recruiting into ongoing research trials. Trinzeck et al have designed and implemented a generic architecture for Patient Recruitment System compatible with most of the currently available German Hospital Information System environnements.

Formal representation of eligibility criteria: The question of the formal representation of eligibility criteria is still an open issue. Several works have been developed or reused formalism for eligibility criteria representation such as CDISC's ASPIRE, Arden Syntax, SAGE, or GELLO or ERGO.

Interoperability and communication: To determine the eligibility status of patient, CRSS has to match clinical data coming from one or several EHR sources to the Clinical trial criteria. This supposes a semantic interoperability and a communication layer between the clinical data sources and the CRSS. To address this issue, the current trend is to combine interoperability standards coming from clinical or research domains (e.g HL7/CDISC), such as the Bridge model.

Methods

Use case addressed by the system: The different steps (S_x) of the MDM process and the system functionalities used all along this process (fig. 1) are : **S1.** A clinical research associate or a principal investigator enter the eligibility criteria of all open clinical trial into the system. This step consist of translating the free text criteria into a computable representation with a graphic user interface. **S2.** Mr. Dupont, a patient having a prostate cancer, consults with Dr. Durant, a urologist. **S3.** Dr Durant schedules a MDM to decide the best therapeutic strategy for his patient. Dr. Durant fills up a MDM form provided by a regional EHR. This MDM report contains the most important medical information about Mr. Dupont, that the MD Team will use during the meeting to take the decision. **S4.** Before the meeting, all the MDM reports are sent through a web service to the system. The system analyses, for each MDM report, all the criteria of all open clinical Trials. **S6.** As output, the system provides an eligibility report for each patient, containing all proposed and rejected clinical trials with the reasoning of the system decision. **S7.** During the meeting, both MDM report and eligibility report are displayed to the multidisciplinary team. The case of Mr. Durant is discussed and a decision is taken by the team. The decision is added to the MDM report. **S8 :** Either, the therapeutic decision is a standard treatment. The MDM report with the decision is then sent to Dr. Durant who informs Mr. Dupont. The patient then either accepts or rejects the proposal. **S9 :** Or according the conclusion of the eligibility criteria report provided by the system, the multidisciplinary team can propose to Mr. Durant to be enrolled into a clinical trial. **S10 :** In this case, a Clinical Research Associate (CRA) is warned after the MDM by a notification. The CRA is in charge to verify details the eligibility of Mr. Durant to. Eventually, Mr. Dupont is contacted, to sign the CT consent. Most of the time, specific exams are scheduled (**S11**) to determine whether the patient is eligible or not. Mr. Dupont is finally included (**S12**).

System Architecture

Patient data source and communication components:

The system was tested with the oncology EHR of the regional oncology network of Brittany. The EHR communicates to the system through secured webs services. A semantic interoperability framework was defined based on the HL7/CDAr2 Level 3 standards and more specifically on HL7 Care Provision/R-MIM Care Record (DSTU) and HL7v3 Standard Transport Specification, Web services Profile, Simple Object Access Protocol version 1.2 (SOAP 1.2). HL7 R-MIM was constrained and only useful classes for the project were implemented. From this refinement, we have produced a set of HL7 templates to formalise and encode the set of data elements of the MDM reports. To encode both the patient data and the CT criteria, the National Cancer Terminology Thesaurus (NCIT) was chosen as reference ontology/terminology. For each data elements of the MDM report, entities and related value were manually mapped to NCIT concepts. In case of missing concepts, we have enriched the ontology by selecting codes and labels from other reference terminologies (e.g. SNOMED CT, LOINC, etc). The communication between the patient data source and the system was performed by a securitized web service. At the message reception, a virtual Medical Record (vMR) was populated by the system with the transmitted patient data (vMR is a generic HL7 data model dedicated to decision support systems).

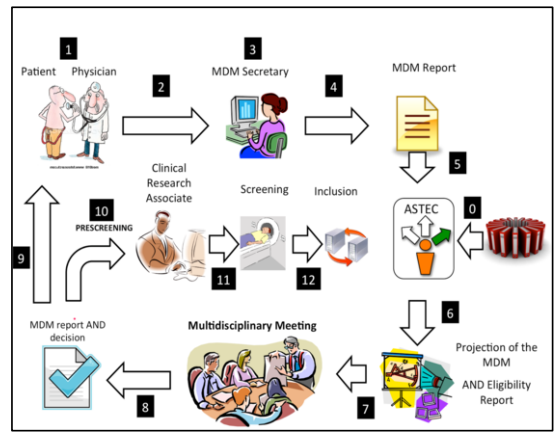


Figure 1 – MultiDisciplinary Meeting workflow

Eligibility and patient data representation: EC refer to complex conditions that a patient needs to fulfill to be included or excluded from a given clinical trial. All clinical trials can be seen as a concatenation of a set of inclusion criteria and a set of exclusion criteria. Patients need to meet the inclusion criteria and avoid the exclusion criteria in order for them to be considered in the clinical trial. Patient data as well as eligibility criteria were formalized relying on OWL models. Namely, patients are defined as a set of *entities*, to which several properties can be associated. Every *property* has a unique *value* that represents its state.

Formalization and processing of the eligible criteria: EC are written by physicians so the terminology used can be subject to interpretation and involve several simple predicates. Thus, a specific criterion can be complex and depend on several *entities.* Complex criteria such as the GETUG 14 eligibility criterion "Cancer in intermediate prognostic group", were decomposed into a set of *atomic criterion*, i.e. a simple predicate on a unique OWL triplet (e, p, v) and a logic AND/OR relationship. As example, Fig. 2 illustrates how this criterion can be broken down to simple predicates (right side of the graph, i.e., leaves of the graph) and a set of Boolean operations.

A patient is included in a CT if his root criteria score is equal to 1. Only patients that met all inclusion criteria and met none of the exclusion criteria received a score of 1.

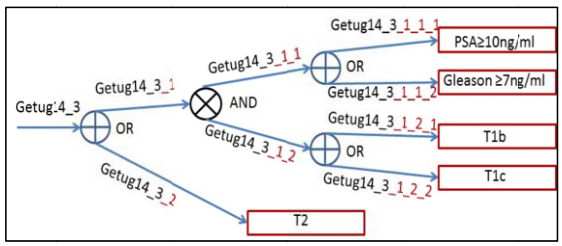


Figure 2-…

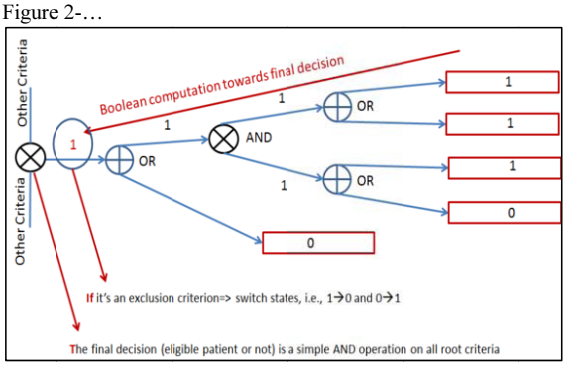


Figure 2- Criterion representation

Inference method: The inference method used in the system has been explained in details in our previous paper [8]. The core of the system is a simple ontology that glues together the components: it imports NCIT ontology and the ontologies that define SWRL built-ins and the classes they could require. This glue ontology defines the classes that are used to represent patient data and the EC in CTs. EC expressed in SWRL are inferred with patient data by an inference engine (JESS). The information granularity is most of the time different between patient data and the EC. This issue is addressed by Jess - subsumption reasoning capability.

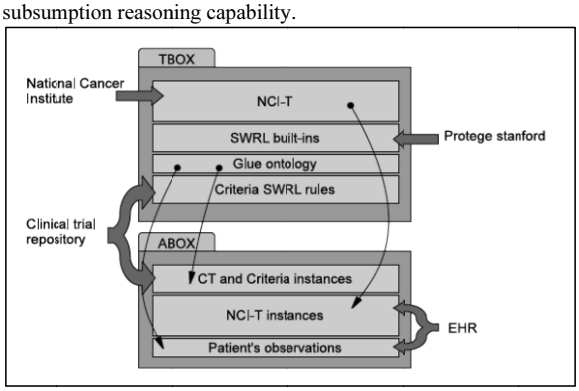


Figure 3 – System design

Figure 3 shows the general architecture: TBox contains the glue ontology that imports NCIT and the SWRL ontologies. The SWRL rules are definitions, and therefore are part of the TBox. The ABox contains the instances of NCIT classes, which are linked by observations, and the instances of the EC.

The eligibility report: A clinical trial has a list of arguments in favour and one against enrolment of the patient. The list in favour is filled by inclusion criteria, while the list against by exclusion criteria. Once the criteria rules are run, the instances of the criteria will have the property *supported_by* either filled with one or more observations, or empty. Inclusion criteria with the property filled are supported, and they in turn support the clinical trial. Exclusion criteria with the *supported_by* property filled are arguments against the enrolling of the patient to the clinical trial. At this point, rules specific to CTs are run to aggregate the results of the criteria. All inclusion criteria of a CT need to be verified: there must be, for each CT, a rule stating that the conjunction of all inclusion criteria must have at least one supporting argument. It is also possible to trace the exclusionary criteria to an observation that fits it. However, extracting the CTs that are supported without arguments against it requires to reason with the closed world assumption: according to the open world assumption, the empty list of arguments against enrolment is considered ignorance, and cannot be used to infer that the there are no arguments against enrolment. The last step needs to be performed outside the ontology. An external program obtains the list of all clinical trials and selects those that are both supported and have no arguments against.

Evaluation of the system

Clinical trials for the evaluation of ASTEC

To evaluate our proposed system, we performed retroactive chart review of all patients in every MDMs from October 2008 to march 2009. Only a selection of CTs recruiting at the CEM during the study period was chosen for the ASTEC evaluation. The study focused on prostate cancer, which is the most common cancer in urology and the main topic of discussions at the urology MDM and of many clinical trials. All criteria of the selected CTs were considered for the study, but required some interpretations to be coded in the system. Some high level criteria were rather vague (e.g. "life expectancy >=10"). Moreover such an item is not present explicitly in the multidisciplinary reports (and probably rarely in most of EHRs). In this case, the Clinical Research Assistant (CRA) in charge of the system setting in accordance with the principal investigator, encoded a set of sub-criteria, that could be match with the information present in the MDM report (to continue with life expectancy ≥ 10, it was encoded in ASTEC as no hepatic failure AND no renal failure … AND Age < 75). Eventually, the aggregating algorithm described above, was used to recompose the initial eligibility criteria and determine the final status of eligibility according to the availability of the real data of the MDM reports.

Automatic screening evaluation and Statistical analysis:

We compared screening performed by the ASTEC system with screening performed by two kinds of human decisions:

- The "MDM Decisions", i.e. the real-time decisions taken during MDM, which are finally the real screening decisions;
- The "Reviewer Decisions", i.e. the decisions taken after review by the CRA and the Principal Investigator (PI) of each study. For that, the overall MD reports were reviewed, manually and independently from the MDM decisions, to propose eligible patients to the RCT. Standard performance metrics are used to assess information retrieval: true and false positives, true and false negatives, recall, precision, f1-measure and overall error rate.

$$precision = \frac{tp}{tp + fp} \quad recall = \frac{tp}{tp + fn}$$

$$f1 - measure = 2.\frac{precision.recall}{precision + recall}$$

Results

During the study period, 23 Multidisciplinary meetings took place with about 285 medical reports. Four ongoing CTs (GETUG 14, 15, 16, 17) concerning prostate cancer were recruiting at the CEM during the study period. Only results from the GETUG 14 are presented, which was the CT with the most eligible patients.

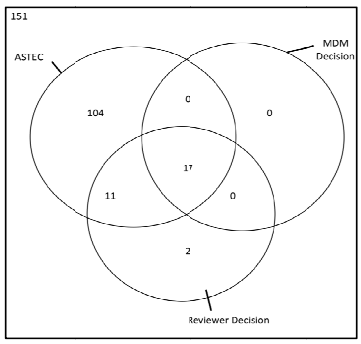


Figure 4 - Venn diagram of the ASTEC decisions, the Reviewer decisions and the MDM decisions

We considered the screening reviewer decisions as gold standard, while the MDM decisions reflected the screening performed in real world by physicians.

As shown in Figure 4, the 3 screening methods all excluded 151 patients. The MDM decisions selected 17 patients eligible to GETUG 14, while the ASTEC system selected 132 patients and the reviewer decisions selected 30 patients eligible to GETUG 14. Finally, the automatic patient screening by the ASTEC system collected 11 additional patients from the selection of the MDM team, and only 2 patients eligible to GETUG 14 were missed.

Table 1–. Performance metrics of the ASTEC system

	MDM Decisions (1)	**Reviewer Decisions (2)**
n total	285	285
n selected patients	17	30
ASTEC classification results with (1) or (2) as reference		
True positives	17	28
False positives	115	104
True negatives	153	151
False negatives	0	2
Precision	13%	21%
Recall	100%	93%
Error Rate	40%	37%
F1-measure	23%	34%

The ASTEC system had a 21% precision rate and 93% recall rate, and an error rate equal to 37% (**Error! Reference source not found.**).

Error! Reference source not found. presents if each GETUG 14 eligible criterion was verified, not verified, or unknown (when data is missing) for all the screened patients. For instance, only the histological confirmation of the prostate cancer was collected in all the MDM reports.

Table 2– Status of the GETUG 14 eligible criteria

Criteria	**Verified**	**Not verified**	**Unknown**
Histologically confirmed prostate cancer	285 (100%)	0	0
Intermediate-risk cancer	57 (20%)	41 (14%)	187 (66%)
No lymph node invasion	20 (7%)	4 (1%)	261 (92%)
No metastatic disease	14 (5%)	5 (2%)	266 (93%)
PSA < 30 ng/mL	261 (92%)	10 (4%)	14 (5%)
No history of invasive cancer	1 (0.4%)	0	284 (99.6%)
Life expectancy ≥ 10 years	89 (31%)	34 (12%)	162 (57%)
No prior pelvic radiotherapy	27 (9.5%)	2 (0.7%)	256 (89.8%)
No radical prostatectomy (and TURP < 3 months)	34 (11.9%)	5 (1.8%)	246 (86.3%)
No prior hormonal therapy and castration	31 (10.9%)	0	254 (89.1%)
Other exclusion criteria (x5)	171 (12%)	379 (26.6%)	875 (61.4%)

Discussion

Semantic interoperability

We developed an interoperability framework using HL7 standards and the ASTEC ontology (based on the NCIT) to encode the data elements. This approach ensures through a service oriented architecture, an effective and secure communication between the oncology EHR and the CRSS. As far as we know, this is the first CRSS project using this kind of interoperability approach. Despite the fact we tested our system on a single commercial EHR, this architecture will help to connect the system to any kind of data sources using the same communication framework.

Reasoning on structured and coded data

A new contribution of ASTEC is that it demonstrates how it is possible to represent eligibility criteria of clinical trials using SWRL on top of a large domain ontology. Some of the criteria were directly translated into SWRL. Others require more thought, especially these which involve temporal reasoning.

In our project, time of processing is not a stringent requirement: the matching of the patients to the available clinical trials is done offline, before the MDM. The slowest step in the overall procedure is loading NCIT ontology. On a dual core machine, with 8Gb of memory, the process takes over 100 seconds and 2Gb of memory. Importing data into the ontology is nearly instantaneous: we load one patient at the time, and only the clinical trials currently active in the hospital are loaded. The next bottleneck is the conversion of the ontology and of the SWRL rules into Jess, which takes on average 10 seconds. The actual running of the engine takes less than a third of a second. Compared to X[1], we use a much smaller ontology (SNOMED CT is over a million classes, while NCIT is only 75000). Loading the background ontology is performed once.

Integration of ASTEC into the business process of MDM

We have designed ASTEC to be seamlessly integrated into the workflow of MDM. However, in this paper, since we have used a retrospective dataset, we don't provide here any evidence of the user acceptance. This is the object of a current study. Regardless, we can give some arguments in favor. As the data processing occurs before the MDM, the eligibility report is available immediately when the case is discussed by the medical MD Team. The system requires very few interactions from the users. During the MDM, the discussions of the experts are extremely short and efficient, so we don't believe that in case of error or misclassification by ASTEC, the users will have the time to modify and rerun the system on line. ASTEC is not intrusive in the decision process, and it behaves as a reminder system, giving for each patient, a short and clear argumentation in favor or against its selection into a CT.

Evaluation of the performance

The results shows that most of the patients eligible to GETUG 14 were selected by the ASTEC system (28/30 patients). Despite numerous false positives (104 patients), the ASTEC system allowed to rightly exclude 151 patients automatically. The 2 patients missed by the ASTEC system, was also missed by the MDM team.

However, we recognise some limitations. Despite considering 4 different CTs at the beginning of the study, only the results for GETUG 14 are presented here. Indeed, 0, 6 and 1 patients were eligible to GETUG 15, 16 and 17 respectively, while 30 eligible patients were eligible for GETUG 14. This led us to solely discuss ASTEC results on a single clinical trial.

Moreover, GETUG is focused on prostate cancer which is one specific domain in oncology.

The question of the scalability of our system to a broader medical domain is still open. But we can assume that the system will be robust at least for cancer domain. Firstly, because we used quite generic interoperability and reasoning methods. Secondly, because the system is based on a recognized and comprehensive terminology in this domain. And finally, and because the process of MDM is quite similar for all cancers.

Dealing with low data quality and missing data

It is worth noting that decomposing patients' conditions into a set of entities is of course challenging as many entities can be only considered at the physician's discretion. We showed in a preliminary study [9] that a CDSS as the ASTEC system must deal with a high level of missing data and unknown information for the majority of the eligibility criteria. Indeed, we demonstrate the existence of implicit information (i.e either data not mentioned into MDM reports, or other eligibility criteria than those mentioned by the protocol) used by physicians to perform the screening task. For instance, when a given condition is not present (e.g., heart condition or failure), it is usually omitted and not explicitly set as such into medical record. A physician can overcome this lack of information, but it is hardly the case when it comes to automatic systems. Moreover, we found in the preliminary study a low rate of data quality from the oncology EHR. Initially, we had assumed that multidisciplinary reports were forensic documents, containing mandatory, explicit and comprehensive information. Indeed, the critical decision of the multidisciplinary team is supposed to rely on this set of information. Actually, it turned out that multidisciplinary reports were just a short summary of the main information. They did not cover all the aspects of the patient condition, such as psychological or psychiatric conditions. We also noted that some of very important data which were supposed to be coded by the clinician in the dedicated structured fields of the MDM form, were present but in the free text fields of the reports. So, these data were unusable by the ASTEC system. For instance, in some multidisciplinary reports, TNM staging was not filled out in the dedicated field of the report and was mentioned in the free text but in an implicit way (e.g. "malign tumor of the both sides of the prostate without extension" which implicitly, stands for a T2 stage). It was unlikely for ASTEC, the multidisiplinary team or the reviewer to deduct at a glance the cancer stage and account for this information during decision making.

Such factors explain the main reasons for discrepancies between the system decision and the human decisions, especially based on false positive screened patients by the system. It is nevertheless considered as a decent approximation, one that can enable an efficient pre-selection and help physicians focus on a subset of promising patients.

To address the issue of missing data, we have developed and tested in a related work [8], an OWL model of clinical trial eligibility criteria compatible with partially-known information. In this work, we proposed an OWL design pattern for modeling the eligibility criteria based on the open world assumption. Our approach successfully distinguished clinical trials for which the patient is eligible, clinical trials for which we know that the patient is not eligible and clinical trials for which the patient may be eligible provided that further pieces of information (which we can identify) can be obtained. The OWL study was evaluated on the same clinical trial and on the same set of patients. The results were similar to the ones reported here, with 149 patients potentially eligible patients (132 in the present study). The difference of performance lies in a simpler data extraction process for the OWL model, which does not affect reasoning. Once the data are extracted from the patient's records, the determination of the patient's eligibility requires a reasoning framework capable of handling both (i) the difference of precision between the data and the criteria, and (ii) the pervasive incompleteness of data. Together, these two studies demonstrate the feasibility of such a task.

Conclusion

Clinical trials are required for the evaluation of medical treatments. Their weakness lies in the difficulty of recruiting enough patients in order to make them statistically meaningful. In this paper we have presented an approach based on OWL and SWRL that addresses the problem of recruitment of patients. The evaluation showed that it is possible to represent the great majority of criteria, and that the system detected most of the patients eligible to the trial and eliminate most of the False negative. The false positives were essentially due to missing data coming from the EHR.

Acknowledgments

The French National Agency of Research (ANR) under the TecSan program 2008 funded the ASTEC project (ASTEC n°ANR-08-003). The authors would like to especially thank Sahar Bayat, Elisabeth Le Prisé, Jean-François Laurent, Nicolas Garcelon, Renaud De Crovoisier for their valuable help.

References

[1] M. Cuggia, P. Besana, and D. Glasspool, "Comparing semi-automatic systems for recruitment of patients to clinical trials," *Int J Med Inform*, vol. 80, no. 6, pp. 371–388, Jun. 2011.

[2] James J Cimino and Elaine J Ayres. The clinical research data repository of the us national institutes of health. Studies in health technology and informatics, 160(Pt2):1299–1303, 2010.

[3] G. D. Moor, et al "Using Electronic Health Records for Clinical Research: the Case of the EHR4CR Project," *J Biomed Inform*, Oct. 2014.

[4] H. J. Lowe, T. A. Ferris, P. M. Hernandez, and S. C. Weber, "STRIDE – An Integrated Standards-Based Translational Research Informatics Platform," *AMIA Annu Symp Proc*, vol. 2009, pp. 391–395, 2009.

[5] H Stenzhorn, et al. The ObTiMA system - ontology-based managing of clinical trials. Studies in health technology and informatics, 160(Pt 2):1090–1094, 2010.

[6] P Vivek, et al"Cancer Multidisciplinary Team Meetings: Evidence, Challenges, and the Role of Clinical Decision Support Technology." *International Journal of Breast Cancer* 2011 (12 2011): 1-7. doi:10.4061/2011/831605.

[7] B. Trinczek, « Design and multicentric implementation of a generic software architecture for patient recruitment systems re-using existing HIS tools and routine patient data », *Appl Clin Inform*, vol. 5, n° 1, p. 264283, 2014.

[8] O. Dameron, et al. , "OWL model of clinical trial eligibility criteria compatible with partially-known information," *J Biomed Semantics*, vol. 4, no. 1, p. 17, 2013.

[9] B Campillo-Gimenez, et al. Improving the pre-screening of eligible patients in order to increase enrollment in cancer clinical trials. Trials 2015- accepted

Author Correspondance

Marc CUGGIA (marc.cuggia@univ-rennes1.fr)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-569

Assessing the Collective Population Representativeness of Related Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES

Zhe He[a], Shuang Wang[b], Elhaam Borhanian[a], Chunhua Weng[a]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*
[b] *Department of Biostatistics, Columbia University, New York, NY, USA*

Abstract

Randomized controlled trials generate high-quality medical evidence. However, the use of unjustified inclusion/exclusion criteria may compromise the external validity of a study. We have introduced a method to assess the population representativeness of related clinical trials using electronic health record (EHR) data. As EHR data may not perfectly represent the real-world patient population, in this work, we further validated the method and its results using the National Health and Nutrition Examination Survey (NHANES) data. We visualized and quantified the differences in the distributions of age, HbA1c, and BMI among the target population of Type 2 diabetes trials, diabetics in NHANES databases, and a convenience sample of patients enrolled in selected Type 2 diabetes trials. The results are consistent with the previous study.

Keywords:

Clinical Trial; Patient Selection; Selection Bias.

Introduction

Randomized controlled trials are the gold standard for producing high-quality medical evidence. However, they may suffer from delayed enrollment and lack of population representativeness, resulting in compromised generalizability to the real-world patients to whom the results or findings of a trial are usually applied. To assess the external validity of a trial, researchers often compared the study population of its enrolled patients and the patient population with certain medical conditions, e.g., breast cancer [1] or major depression [2]. Our previous work showed that many clinical trials, especially those on the same medical condition, use similar or identical eligibility criteria [3]. Therefore, we hypothesize that the generalizability issue might be at not only the individual trial level but also the community level of the entire clinical trial enterprise.

As a step to advance the field for generalizability assessment, we previously proposed a method to compare patient populations in electronic health records (EHRs) with aggregated clinical trial target populations that characterize the patients who can be recruited in a set of trials according to their inclusion and exclusion criteria [4]. This method also introduced *Generalizability Index for Study Traits (GIST)* for quantifying the population representativeness of clinical trials to the general patient population. This approach is advantageous over existing methods [1, 2] in that generalizability assessment can be performed proactively during design. Using this method, we found that Type 2 diabetes studies are more generalizable with respect to age than they are with respect to HbA1c. However, one of the limitations of the study is the potential bias in EHR data towards certain population subgroups [5].

The available public datasets offer great opportunity to further validate the method and the results generated from EHR data. The National Health and Nutrition Examination Survey (NHANES) is a continuous cross-sectional health survey conducted by the National Center for Health Statistics of CDC [6]. It evaluates a stratified multistage probability sample of the non-institutionalized population of the United States. The survey samples are first interviewed at home, followed by a physical and laboratory test in a mobile examination center. Its rigorous quality control standards ensure national representativeness and high-quality data collection. NHANES data have facilitated various translational bioinformatics research. Chen *et al.*, for example, have built a predictive aging model of adolescent development [7]. Bays *et al.* have analyzed connections between body mass index and metabolic diseases [8]. These promising results have propelled our use of NHANES data to support our population-based study. In this paper, we assess the collective population representativeness of multiple related Type 2 diabetes mellitus (T2DM) trials through pair-wise comparison between the patient population derived from the NHANES data, the target population derived from clinical trial summaries, and a convenience sample of study population derived from the results of selected trials in ClinicalTrials.gov. We hypothesize that NHANES data can be used to assess the population representativeness of trials and to validate the results of our previous study using EHR data [4].

Methods

Figure 1 shows the workflow of this study. We first identified frequently used quantitative eligibility features (i.e., with a permissible value range) in T2DM trials between 2003 and 2012.

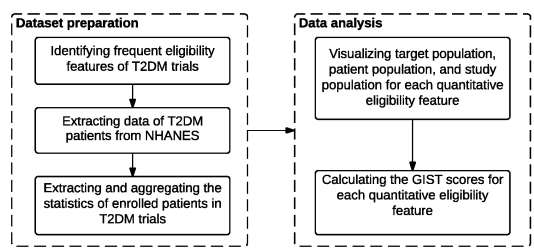


Figure 1– The workflow of the study

In the NHANES data of continuous survey years 2003 to 2012, we identified sample subjects with T2DM and extracted interview and laboratory test results relevant to this work. We

also extracted results of selected T2DM clinical trials between 2003 and 2012 from ClinicalTrials.gov and used these data to profile a convenience sample of the study population. We visualized and analyzed the differences among the T2DM patient population, the target populations of T2DM trials, and a convenience sample of study populations of enrolled patients in T2DM trials, all of which were profiled by one eligibility feature at a time. Then we calculated the GIST scores [4] for each eligibility feature.

Dataset Preparation

Processing clinical trial summaries

We have built and published a database COMPACT [9], which stores parsed eligibility criteria and trial descriptors of 162,586 clinical study summaries in ClinicalTrials.gov downloaded on March 18, 2014. All the trials were indexed by medical conditions using the API provided by ClinicalTrials.gov. From COMPACT, we retrieved 2,731 interventional studies on T2DM with a start date between 01/2003 and 12/2012. Age, HbA1c, and body mass index (BMI) were identified as three most frequently used quantitative eligibility features that appear in 2,702 (99.0%), 1,463 (53.6%), and 1,274 (46.6%) of the trials, respectively. We formed one trial set for each of the three features and included them for analyzing the distribution of trials over age, HbA1c, and BMI, respectively.

Processing NHANES data

We extracted relevant interview data and laboratory measurements for each NHANES sample during 2003 – 2012. The data elements included *SEQN* (the unique identifier of a sample), *gender*, *age*, *race/ethnicity*, *BMI*, *Glycohemoglobin (HbA1c) values*, *two-year sample weights for the interview and mobile examination portion of the study*, *two year cycle*, and interview questions "*Were you told to have diabetes by a doctor or a health professional*", "*Age when first you were told to have diabetes*", and "*Are you taking insulin?*" The laboratory methodology to measure Glycohemoglobin value of NHANES sample subjects can be found at [10].

In NHANES, we identified 3,304 diabetic sample subjects who were told to have diabetes by a doctor or a health professional and had an HbA1c measurement [11]. Since NHANES participants were not asked to report the type of diabetes (type 1 or type 2) they were diagnosed with, we employed a method used by Dodd *et al.* [11] to further exclude 222 samples with Type 1 diabetes who were (1) first diagnosed with diabetes before age 30; and (2) taking insulin. The rationale is that as one grows older, his/her lifestyle (e.g., dietary habits) will play a more important factor in developing T2DM; thus, if a person is diagnosed with diabetes at a young age and takes only insulin, it is likely that s/he is with Type 1 diabetes.

Out of the 3,082 T2DM samples, 2,695 had no missing values for age, HbA1c, and BMI. Because the number of samples with missing data (3,082 - 2,695) exceeded 10% of the total number of T2DM samples, we used two categorical variables gender and race/ethnicity that every sample of NHANES has to assess the representativeness of these 2,695 samples for all 3,082 T2DM samples. With respect to gender and race/ethnicity, we used the chi-square test to test the statistically significant differences between any pair of the following three sets of samples: (1) all 3,082 samples, (2) 2,695 samples with age, HbA1c and BMI values, and (3) 387 samples with missing values. No statistically significant pair-wise difference was found between any pair of samples (all *p-values* > 0.05). Further, we used the two-sample *t*-test to test the pair-wise differences for HbA1c, age, BMI. No statistically significant difference was found (all *p-values* > 0.05). Therefore, we concluded that the 2,695 samples that had complete age, HbA1c, and age values is a representative sample of all the T2DM samples in NHANES and we used this sample as the patient cohort of this study.

To account for complex survey design (e.g., oversampling), non-response, and post-stratification, NHANES assigned each sample subject a sample weight, which represents the number of people in the U.S. national population that a specific sample can represent. When a sample is weighted, it is representative of the U.S. civilian non-institutionalized Census population [12]. NHANES publishes survey data once every two years. In this study, we combined data of five two-year cycles from 2003 to 2012. Following the analytical guideline of NHANES [13], we used WTMEC2YR as the sample weight and calculated the ten-year sample weight WTMEC10YR (1/5 * WTMEC2YR). After taking the ten-year sample weight into account, these 2,695 samples can thus represent 15,575,484 T2DM patients in the U.S. national population. More importantly, the distribution of patients with sample weight can represent the real distribution of the U.S. national population. In this paper, all the subsequent analyses were performed after taking sample weights into account. Table 1 shows the baseline characteristics of the T2DM patient cohort of the study.

Table 1– Baseline characteristics of patient cohort

Characteristics	Number
Sample, n	2,695
Population, n	15,575,484
Age (mean ± SD)	60.6 ± 13.3
Gender, male%	48.2
Race/Ethnicity	
Mexican American, %	8.5
Other Hispanic, %	5.5
Non-Hispanic White, %	61.8
Non-Hispanic Black, %	16.0
Other races, %	8.2
HbA1c, % (mean ± SD)	7.2 ± 1.7
BMI, kg/m^2 (mean ± SD)	32.9 ± 7.6

Processing clinical trial results

To profile the study populations, we parsed the baseline characteristics of enrolled patients for the T2DM studies that published results in ClinicalTrials.gov. Out of 2,731 T2DM trials between 2003 and 2012, only 531 reported their baseline characteristics of enrolled participants in ClinicalTrials.gov. The numbers of trials that reported mean and standard deviation (SD) of age, HbA1c, and BMI of their enrolled participants are 389, 137, and 108, respectively. Since only a small portion of trials reported results in ClinicalTrials.gov, the study population included in this analysis represents merely a convenience sample of the study population. For each study, we extracted participant counts, mean and SD for age, HbA1c, and BMI. We aggregated the mean and SD for each feature using the following formulas (adapted from [14]), where T is the number of studies,

$$\boldsymbol{Weighted\_mean} = \frac{\sum_{i=1}^{T}(mean_i * number\_participants_i)}{\sum_{i=1}^{T} number\_participants_i} \quad (1)$$

$$\boldsymbol{Weighted\_SD} = \sqrt{\frac{\sum_{i=1}^{T}(SD_i^2 * (number\_participants_i - 1))}{\sum_{i=1}^{T}(number\_participants_i - 1)}} \quad (2)$$

Visualizing populations

Effective visualization of target populations, patient population, and study population over a quantitative trait allows interested viewers to easily discern the differences among them. To profile the target populations of T2DM

studies represented by a quantitative eligibility feature, we plotted the distribution of trials over a quantitative feature (i.e., age, HbA1c, and BMI). This distribution shows the number of trials that recruit patients with a certain value, which can reveal systematically excluded or overly included value ranges. The patient population of T2DM was presented by the distributions of several patients' features. The study population was similarly plotted using Gaussian distributions with weighted means and SDs of several features.

For each feature, we plotted the distributions of patient population, target population, and study population in the same figure. We employed 2nd degree polynomial local weighted regression fitting with a span of 20% to smooth the curves [15].

Quantitative metric of population representativess

To quantify the population representativeness of studies for a single quantitative feature, such as age, HbA1c, and BMI, we calculated the GIST scores [4]. GIST is the sum across all consecutive non-overlapping value intervals of the percentage of studies that recruit patients in that interval, multiplied by the percentage of patients observed in that interval:

$$GIST = \sum_{i=1}^{N} \frac{\sum_{j=1}^{T} I([i_{low}, i_{high}] \subset w_j)}{T} * \frac{\sum_{k=1}^{P} I(i_{low} \le y_k < i_{high})}{P} \quad (3)$$

where N is the number of distinct value intervals of the quantitative feature, T is the number of trials, P is the number of patients, w_i is the inclusion value interval of the quantitative feature for the j^{th} study, such that indicator I can be defined as j^{th} study interval subsumes the i^{th} interval low and high boundary values, and y_k is the observed value of the quantitative feature for the k^{th} patient such that an indicator I can be defined when k^{th} patient has a value of the quantitative feature falls within the i^{th} interval.

The GIST score ranges between 0 and 1 and characterizes the proportion of patients that would be potentially eligible across trials, with 1 being perfectly generalizable and 0 being completely not generalizable. Note that the formula for calculating the GIST score can also be applied to categorical variables, whereby the values are integers.

Results

Visualization of populations

Figure 2 shows the juxtaposition of the distributions of T2DM patients (blue solid curve), T2DM trials (green dot-and-dashed curve), and study samples enrolled in T2DM trials (red dashed curve) over age values. The x-axis represents the age values. The left y-axis represents the percentage of patients of a certain age. The right y-axis represents the percentage of trials that recruit patients of a certain age. Note that both target population (green dot-and-dashed curve) and study population (red dashed curve) use the same value ticks on the right y-axis. From the visualization, we can observe that T2DM trials recruit patients with broad age range (between 18 – 80), whereas most real-world patients are older. Comparing the target population (green dot-and-dashed curve) and study population (red dashed curve), we can see that even though T2DM studies tend to recruit patients of broad age range, most enrolled patients were between 40 to 70 years old.

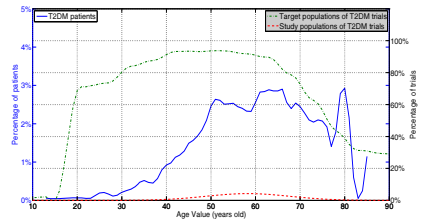


Figure 2– Target population, patient population, and study population of T2DM trials on age values

Figure 3 visualizes the comparison of these three populations over HbA1c values. Between years 2003 and 2012, targeted HbA1c values in T2DM trials are consistently higher than that of the patient population, indicating that T2DM trials tend to recruit sicker diabetic patients. This observation is consistent with the results of our previous study [4]. There is also a noticeable shift between the study population (red dashed curve) and the patient population (blue solid curve), indicating that on average patients enrolled in T2DM trials had a higher HbA1c value than the general T2DM patient population. The curves for study patients and target patients aligned well with similar curve shapes.

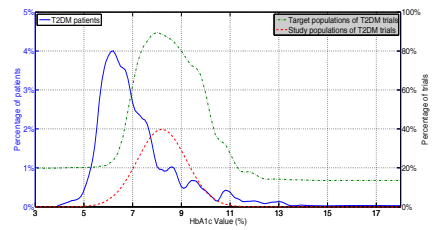


Figure 3– Target population, patient population, and study population of T2DM trials on HbA1c values

Figure 4 is the visualization for BMI. We can see that three curves aligned better than those for age and HbA1c, indicating that T2DM studies may have a better population representativeness with respect to BMI. The concentration of BMI values is accordance with real-world T2DM patients.

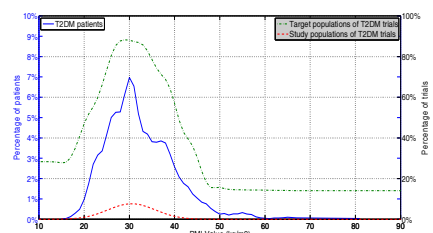


Figure 4– Target population, patient population, and study population of T2DM trials on BMI valuesComparison: study population vs. patient population

Table 2 presents the comparison between the study population (i.e., patients who were enrolled in T2DM trials) and the general patient population. One-sample *t*-test was performed to test the statistical significance of the differences between these two populations for each feature, whose mean value of the general patients was used as hypothetical mean. As shown in the table, statistically significant differences between the study population and the general patient population were observed for all three features. On average, the patients enrolled in T2DM trials are younger, with lower BMI and

higher HbA1c than the general patient population ($p < 0.0001$).

Table 2– Comparison between the convenience sample of the study population and the patient population

Feature (*mean* ± *SD*)	Study population	Patient population	Difference, mean (95% CI)	*P value*
N	198,050	15,575,484	--	--
Age	58.3 ± 9.4	60.6 ± 13.3	– 2.3 (– 2.34 to – 2.26)	< 0.0001
N	62,931	15,575,484	--	--
HbA1c	8.2 ± 1.0	7.2 ± 1.7	1.0 (0.99 to 1.01)	< 0.0001
N	70,678	15,575,484	--	--
BMI	30.5 ± 5.2	32.9 ± 7.6	– 2.4 (– 2.44 to – 2.36)	< 0.0001

GIST scores

The GIST scores were computed for age, HbA1c, and BMI for each phase using fixed-width value intervals of width 1.0, 0.5, and 1.0, respectively (Table 3). Note that large value intervals may exclude some trials; thus, we used these small interval widths to ensure that all the trials were included. To calculate the GIST score correctly, one could also use all the threshold values of a feature in all the trials to divide its value spectrum into bins of varying widths. The overall GIST scores show that T2DM trials have the best population representativeness with regard to age, followed by BMI and then HbA1c. Comparing the results across trials in different phases, we observed that the GIST of age increases from Phase I to Phase III, whereas the GIST of HbA1c decreases from Phase I to Phase III, which confirms our previous study using EHR data [4].

We have provided the interim data and results of this work as supplementary material, which can be accessed at http://is.gd/tbwzj9.

Table 3– GIST scores of age, HbA1c, and BMI for each phase

GIST	Variable		
Phase	Age	HbA1c	BMI
I	0.60 (N=368)	0.55 (N=141)	0.64 (N=204)
II	0.77 (N=517)	0.45 (N=244)	0.71 (N=194)
III	0.87 (N=766)	0.38 (N=438)	0.77 (N=356)
IV	0.80 (N=484)	0.42 (N=306)	0.69 (N=194)
All	0.77 (N=2702)	0.44 (N=1463)	0.69 (N=1274)

Discussion

In this study, we used both visualization and quantitative metric GIST to assess the population representativeness of T2DM trials to the general T2DM patient population. One interesting observation is that even though the curves of BMI (Figure 2) aligned better than those of age (Figure 4) in the visualization, the GIST score of BMI (0.69) is lower than age (0.77), showing that visualization may not be sufficient for comparing population representativeness between variables. Conversely, the GIST score may not be as intuitive as visualization for discerning systematic included or excluded value ranges.

Clinical trials may serve varying purposes. Not all clinical trials need to generalize to a broad patient population. Many trials require a "clean" and specific cohort to ensure internal validity, which is usually achieved by minimizing confounding variables with tightened eligibility criteria. Therefore, the purpose of this study is not to demonstrate the systematic bias in T2DM trials, but to uncover collective design patterns via aggregate analysis of multiple trials. We identified and visualized the significant differences among the aggregated study populations, target populations in trials that investigate the same medical condition, and patient populations that are supposed to benefit from these trials. Our method intends to quantify the differences among these three populations so that they can differentiate them for research in any disease domain and make informed decisions for future trial designs.

NHANES data vs. EHR data

Using NHANES data, we generated results that are consistent with our previous study using EHR data [4]. We further validated the effectiveness of the method for assessing the population representativeness of clinical trials. With the large amount of patient data existing in both EHRs and public databases, it is important to understand their respective strengths and weaknesses for different kinds of analyses. NHANES may be more efficient than EHRs for population-based studies. Firstly, EHRs only contain the data of patients who have paid visits to a hospital, whereas NHANES samples are not limited to patients in hospitals and can represent the U.S. national patient population. Secondly, NHANES data are well structured and readily analyzable, whereas the multi-dimensional data quality problem and largely unstructured data elements of EHRs pose significant impediments for their secondary use. Conversely, EHR data usually contain multiple readings for a variable for a patient, allowing longitudinal analysis of disease progression. Most variables in NHANES have only one reading for each sample. Meanwhile, NHANES may not provide sufficient data for assessing the population representativeness of trials on other medical conditions.

Other works for improving generalizability of trials

To improve the generalizability of clinical trials, Arterburn *et al.* proposed a population-based shared decision-making recruitment approach, which achieved a relatively high ratio of outreach-to-randomized subjects [16]. However, the selection biases with this approach may still remain. Frangakis *et al.* discussed various strategies to calibrate the treatment effects from clinical trials to target populations, including calibration of pre-treatment exclusion criteria and post-treatment measures [17]. However, pre-treatment calibration can only generalize categorical and dichotomous variables but not continuous variables such as age, HbA1c, and BMI. Post-treatment calibration can only generalize components of randomized controlled trials that are not causal effects of treatment [18]. Therefore, these methods may not fully resolve the generalizability issue. We hope that this line of research can help trial designers better balance the tradeoffs between internal validity and external validity during the design phase of a new trial, thereby improving its population representatives.

Limitations and future work

Quite a few limitations are noteworthy when interpreting this study. NHANES uses self-reported medical condition. We borrowed a method from Dodd *et al.* [11] to distinguish between Type 2 and Type 1 diabetes patients, which, in conjunction with the self-reported data, may have led to some misclassification of samples.

Even though the Food and Drug Administration Admendments Act (FDAAA) mandates the reporting of basic summary results on registered or approved product to be submitted to ClinicalTrials.gov, only about 20% of T2DM trials between 2003 and 2012 have submitted results to

ClinicalTrials.gov. Fewer than half of NIH-funded trials published their results in scientific journals within 30 months of trial completion [19]. Since trials with positive results tend to report them, the convenience sample of study populations we used in this work may not be a random sample of the enrolled patients in all the T2DM trials. Moreover, the study population plotted in Gaussian distribution may differ from the histogram of the actual patients. Recently, the US Department of Health and Human Services proposed a new regulation to require public sharing of summary data from clinical trials of the Food and Drug Administration (FDA) regulated drugs and devices regardless of their approval status for marketing [20]. As more and more studies publish their results in ClinicalTrials.gov, a uniform reporting mechanism would significantly help aggregate analysis such as the one we performed in this study. Another frequent concern is that clinical trials retrieved by ClinicalTrials.gov's API may have condition-indexing errors, though we checked a random sample of 100 T2DM trials and found that the accuracy of indexing is acceptable for this condition and this study. The accuracy of our natural language processing tool for parsing complex free-text eligibility criteria needs further improvements.

Different eligibility features may have inherent correlations. For example, impaired fasting glucose is positively correlated with age [21]. In the future, we plan to use multiple features simultaneously to assess the population representativeness of trials.

Conclusions

In this study, we identified and visualized statistically significant differences among the T2DM patient population, the target population of T2DM trials, and a convenience sample of the study population extracted from the results of selected T2DM trials in ClinicalTrials.gov. The consistent results with our previous study [4] have demonstrated the feasibility of using NHANES data to assess the population representativeness of clinical trials. We have further validated the results of our previous study using EHR data [4].

Acknowledgments

We thank Dr. Praveen Chandar for the insightful discussion on this work. This study is sponsored by the National Library of Medicine Grant R01LM009886 (PI: Weng) and the National Center for Advancing Translational Science Award UL1 TR000040 (PI: Ginsberg).

References

[1] van de Water W, Kiderlen M, Bastiaannet E, et al. External validity of a trial comprised of elderly patients with hormone receptor-positive breast cancer. J Natl Cancer Inst. 2014: 106(4):dju051.

[2] Blanco C, Olfson M, Goodwin RD, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 2008: 69(8): 1276-80.

[3] Hao T, Rusanov A, Boland MR, Weng C. Clustering clinical trials with similar eligibility criteria features. J Biomed Inform 2014: 52: 112-20.

[4] Weng C, Li Y, Ryan P, et al. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. Appl Clin Inform 2014: 5(2): 463-79.

[5] Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Med Inform Decis Mak 2014: 14:51.

[6] CDC. National Health and Nutrition Examination Survey Data: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available from: http://www.cdc.gov/nchs/nhanes.htm.

[7] Chen DP, Morgan AA, Butte AJ. Validating pathophysiological models of aging using clinical electronic medical records. J Biomed Inform 2010: 43(3): 358-64.

[8] Bays HE, Chapman RH, Grandy S, Group SI. The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. Int J Clin Pract 2007: 61(5): 737-47.

[9] He Z, Carini S, Hao T, Sim I, Weng C. A Method for Analyzing Commonalities in Clinical Trial Target Populations. AMIA Annual Symp Proc 2014: In press.

[10] Description of Laboratory Methodology to Measure Glycohemoglobin in NHANES [October 19, 2014]. Available from: http://wwwn.cdc.gov/nchs/nhanes/2011-2012/GHB_G.htm.

[11] Dodd AH, Colby MS, Boye KS, Fahlman C, Kim S, Briefel RR. Treatment approach and HbA1c control among US adults with type 2 diabetes: NHANES 1999-2004. Curr Med Res Opin 2009: 25(7): 1605-13.

[12] Key Concepts About Weighting in NHANES [October 2014]. Available from: http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/OverviewKey.htm.

[13] CDC. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999–2010 [September 20, 2014]. Available from: http://www.cdc.gov/nchs/data/series/sr_02/sr02_161.pdf.

[14] Pooled Variance [October 19, 2014]. Available from: http://en.wikipedia.org/wiki/Pooled_variance.

[15] Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 1988: 83: 596-610.

[16] Arterburn D, Flum DR, Westbrook EO, Fuller S, Shea M, Bock SN, et al. A population-based, shared decision-making approach to recruit for a randomized trial of bariatric surgery versus lifestyle for type 2 diabetes. Surg Obes Relat Dis 2013: 9(6): 837-44.

[17] Frangakis C. The calibration of treatment effects from clinical trials to target populations. Clin Trials 2009: 6(2): 136-40.

[18] Hatfield J, Faunce GJ, Soames Job RF. Avoiding confusion surrounding the phrase "correlation does not imply causation.". Teaching of Psychology 2006: 33: 49-51.

[19] Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ 2012: 344: d7292.

[20] Hudson KL, Collins FS. Sharing and Reporting the Results of Clinical Trials. JAMA 2015: 313(4): 355-6.

[21] Cowie CC, Rust KF, Byrd-Holt DD, Eberhardt MS, Flegal KM, Engelgau MM, et al. Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health And Nutrition Examination Survey 1999-2002. Diabetes Care 2006: 29(6): 1263-8.

Address for correspondence

Dr. Chunhua Weng, 622 W 168th Street, PH-20, New York, NY 10032, USA. Email: cw2384@columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-574

Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers

George Hripcsak[a], Jon D. Duke[b], Nigam H. Shah[c], Christian G. Reich[d], Vojtech Huser[e], Martijn J. Schuemie[f,g], Marc A. Suchard[h], Rae Woong Park[i], Ian Chi Kei Wong[f], Peter R. Rijnbeek[j], Johan van der Lei[j], Nicole Pratt[k], G. Niklas Norén[l], Yu-Chuan Li[m], Paul E. Stang[g], David Madigan[n], Patrick B. Ryan[g]

[a] *Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA*
[b] *Regenstrief Institute, Indianapolis, IN, USA*
[c] *Center for Biomedical Informatics Research, Stanford University, CA, USA*
[d] *AstraZeneca PLC, Waltham, MA, USA*
[e] *NIH Clinical Center, Bethesda, MD, USA*
[f] *Centre for Safe Medication Practice and Research, Dept. of Pharmacology and Pharmacy, University of Hong Kong, Hong Kong*
[g] *Janssen Research & Development, LLC, Titusville, NJ, USA*
[h] *Dept. of Biomathematics & Dept. of Human Genetics, David Geffen School of Medicine, Uni. of California, Los Angeles, CA, USA*
[i] *Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea*
[j] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[k] *School of Pharmacy and Medical Sciences, University of South Australia, Australia*
[l] *Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden*
[m] *College of Medical Science and Technology (CoMST), Taipei Medical University, Taipei, Taiwan*
[n] *Department of Statistics, Columbia University, New York, NY, USA*

Abstract

The vision of creating accessible, reliable clinical evidence by accessing the clincial experience of hundreds of millions of patients across the globe is a reality. Observational Health Data Sciences and Informatics (OHDSI) has built on learnings from the Observational Medical Outcomes Partnership to turn methods research and insights into a suite of applications and exploration tools that move the field closer to the ultimate goal of generating evidence about all aspects of healthcare to serve the needs of patients, clinicians and all other decision-makers around the world.

Keywords:

Health Services Research; Databases; Observation.

Introduction

Observational Health Data Sciences and Informatics (OHDSI, pronounced "Odyssey") [1] is an international collaborative whose goal is to create and apply open-source data analytic solutions to a large network of health databases to improve human health and wellbeing. The OHDSI team comprises academics, industry scientists, health care providers, and regulators whose formal mission is to transform medical decision making by creating reliable scientific evidence about disease natural history, healthcare delivery, and the effects of medical interventions through large-scale analysis of observational health databases for population-level estimation and patient-level predictions [2]. Over 90 participants from around the world have joined the collaborative with a vision to access a network of one billion patients to generate evidence about all aspects of healthcare, where patients, clinicians and all other decision-makers around the world use OHDSI tools and evidence every day [3].

Methods

OHDSI grew out of the Observational Medical Outcomes Partnership (OMOP) [4], which was a public-private partnership established in the US to inform the appropriate use of observational healthcare databases for studying the effects of medical products. The five-year project developed new methods in observational research and established an observational research laboratory. At the conclusion of this five-year project, the OMOP research investigators initiated the OHDSI effort. The research laboratory moved to the Reagan-Udall Foundation for the FDA under the Innovation in Medical Evidence Development and Surveillance (IMEDS)

Table 1. Tables in the OMOP Common Data Model V5.0

Model Domain	Table Names
Standardized Clinical Data Tables	PERSON, OBSERVATION_PERIOD, SPECIMEN, DEATH, VISIT_OCCURRENCE, PROCEDURE_OCCURRENCE, DRUG_EXPOSURE, DEVICE_EXPOSURE, CONDITION_OCCURRENCE, MEASUREMENT, NOTE, OBSERVATION, FACT_RELATIONSHIP
Standardized Health System Data Tables	LOCATION, CARE_SITE, PROVIDER
Standardized Health Economics Data Tables	PAYER_PLAN_PERIOD, VISIT_COST, PROCEDURE_COST, DRUG_COST, DEVICE_COST
Standardized Derived Elements	COHORT, COHORT_ATTRIBUTE, DRUG_ERA, DOSE_ERA, CONDITION_ERA

Program [5].

A centerpiece of the OMOP project was the development of the OMOP Common Data Model (CDM) [4] which represents healthcare data from diverse sources in a consistent and standardized way (see Table 1). This CDM is a "strong" information model, in which the encoding and relationships among concepts are explicitly and formally specified. The OHDSI team has adopted and continued maintenance of this model and its associated vocabulary services. OHDSI's overall approach is to create an open network of observational data holders, and require that they translate their data to the OMOP CDM. Each element in the participant's database must be mapped to the approved CDM vocabulary and placed in the data schema. In return, this approach creates a unique opportunity of implementing a number of existing data exploration and evidence generation tools and participating in world-wide studies because any given query can be executed at any site without modification. This enables multicenter, global analyses to be executed rapidly and efficiently using applications or programs developed at a single site.

Data are retained at the participant's site, simplifying patient and business privacy issues. The team previously found that simply merging the databases is likely to give poor answers because of heterogeneity [6]. Instead, analyses are carried out locally and the results transmitted to the coordinating center, where they can be studied on a population level and aggregated as appropriate.

OHDSI operates at several levels: infrastructure, data, methods, applications, and experiments. These levels serve both to support and inform the work of each other to ensure that the infrastructure and products support the mission. Rather than just creating a data network, OHDSI directly integrates researchers who use the network and data scientists who create the algorithms with the use cases for the data network.

The group's guiding principles are that the effort be:

1. Evidence-based, such that OHDSI's scientific research and development are driven by objective, empirical evidence to ensure accuracy and reliability;
2. Practical, going beyond methodological research, but developing applied solutions and generating clinical evidence;
3. Comprehensive, aiming to generate reliable scientific evidence for all interventions and all outcomes;
4. Transparent, such that all work products within OHDSI are Open Source and publicly available, including source code, analysis results, and other evidence generated in all our activities;
5. Inclusive, encouraging active participation from all stakeholders – patients, providers, payers, government, industry, academia – in all phases of research and development; and finally
6. Secure, protecting patient privacy and respecting data

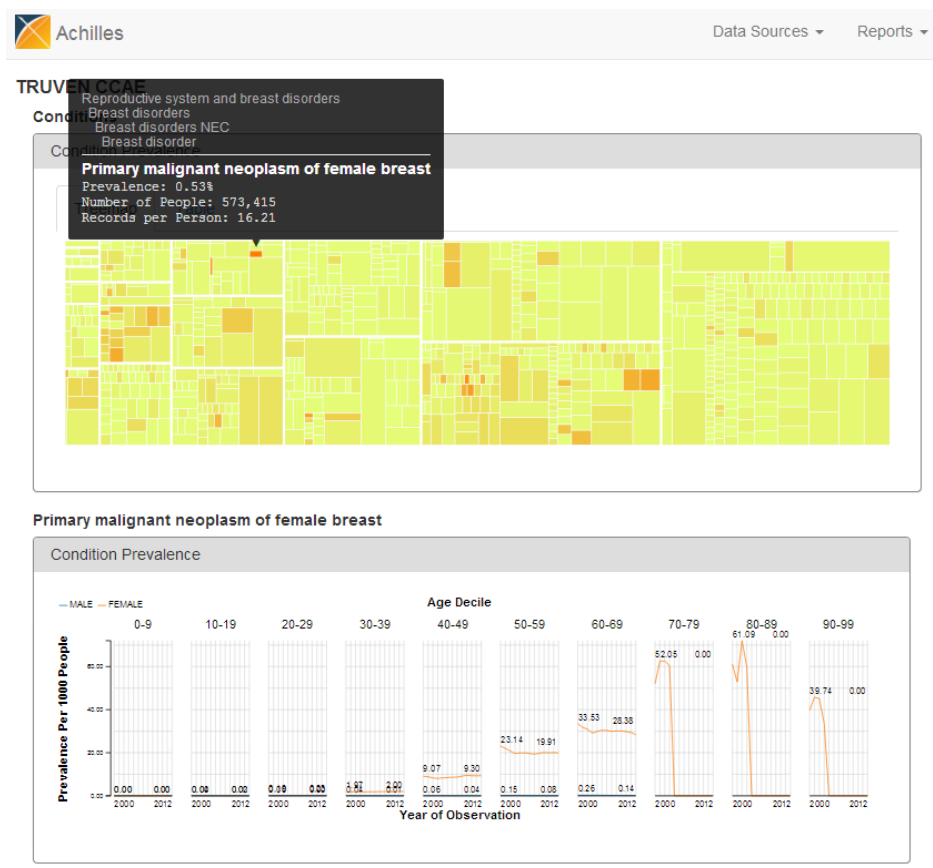


Figure 1. ACHILLES. A tree map (top half) summarizes the content of the database, where the rectangle size indicates prevalence and color indicates average number of records per patient. The bottom half reports prevalence by year, age, and sex.

holder interests at all times.

To achieve the principle of inclusivity, OHDSI is an open collaborative. Anyone who can give time, data, or funding is welcome, and participation in the operation of OHDSI is expected. Currently, participants come from around the world, including the United States, United Kingdom, Netherlands, Sweden, Italy, Korea, Taiwan, Hong Kong, and Australia.

Software Tools

OHDSI is building a suite of tools to faciliate data exploration and evidence generation. The first of these, ACHILLES (Automated Characterization of Health Information at Large-scale Longitudinal Exploration System), is a browser-based visualization tool for reviewing a clinical database based on pre-extracted summary statistics from datasets in OMOP CDM format. ACHILLES enables characterization, quality assessment, and visualization of observational data; and provides users with an interactive, exploratory framework to assess patient demographics and the prevalence of all conditions, drugs, procedures, and observations stored in the dataset. The ACHILLES application and source code is available in the public domain [6] and a demonstration is hosted on the OHDSI Web site [7].

ACHILLES has two main components. The first component is implemented as an R package and runs securely within an organization's local environment without disclosing any patient identifiable information. The R package generates summary statistics that describe the quality and content of the patient-level observational health database and provides features to perform a simple review or bulk export of the summary statistics in JSON data files. The second component of ACHILLES is implemented as an HTML5 / JavaScript website with a series of interactive reports that allow exploration and visualization of generated summary statistics. Summary statistics from multiple databases can be made available from a single installation of the ACHILLES website.

Data owners have used ACHILLES to assess the quality of their database, looking for gaps that may signify upload errors. Other investigators have used ACHILLES to do an initial assessment of whether the database is likely to hold a sufficient number of cases of interest to be worth investigating further. Figure 1 shows a typical visualization for a researcher interested in primary malignant neoplasm of the female breast. As shown, ACHILLES displays the prevalence of the condition, the depth of data on those patients, the age distribution, the sex distribution (less relevant here), and the time of observation. Other current ACHILLES views show the temporal characteristics of the data (e.g., prevalence by month), quality reports, and treemaps of subsets of data where each box's size and color represent different database metrics. Its data quality reports help the data owner curate the database and allow outsiders to review supported aspects of data quality.

Additional OHDSI tools are in development. HERMES (Health Entity Relationship and Metadata Exploration System) is a web-based vocabulary browsing tool with the ability to search for a term and explore related concepts. PLATO (Patient-Level Assessment of Treatment Outcomes) provides predictive models that assess probability of a patient experiencing any outcome following initiation of any intervention, given his or her personal medical history. For example, a patient could enter his gender, age, primary diagnosis, and medication to check the prevalence of side effects to the medication. HERACLES (Health Enterprise Resource and Care Learning Exploration System) helps the user to build and explore cohorts to assess a specific clinical population across a wide-variety of clinical dimensions, including specialized analytics for performing clinical quality metrics. And HOMER (Health Outcomes and Medical Effectiveness Research) enables risk identification and comparative effectiveness studies, with real-time exploration of the effects of medical products. HOMER supports exploratory analyses for a wide variety of dimensions that serve as evidence for or against causality, such as the

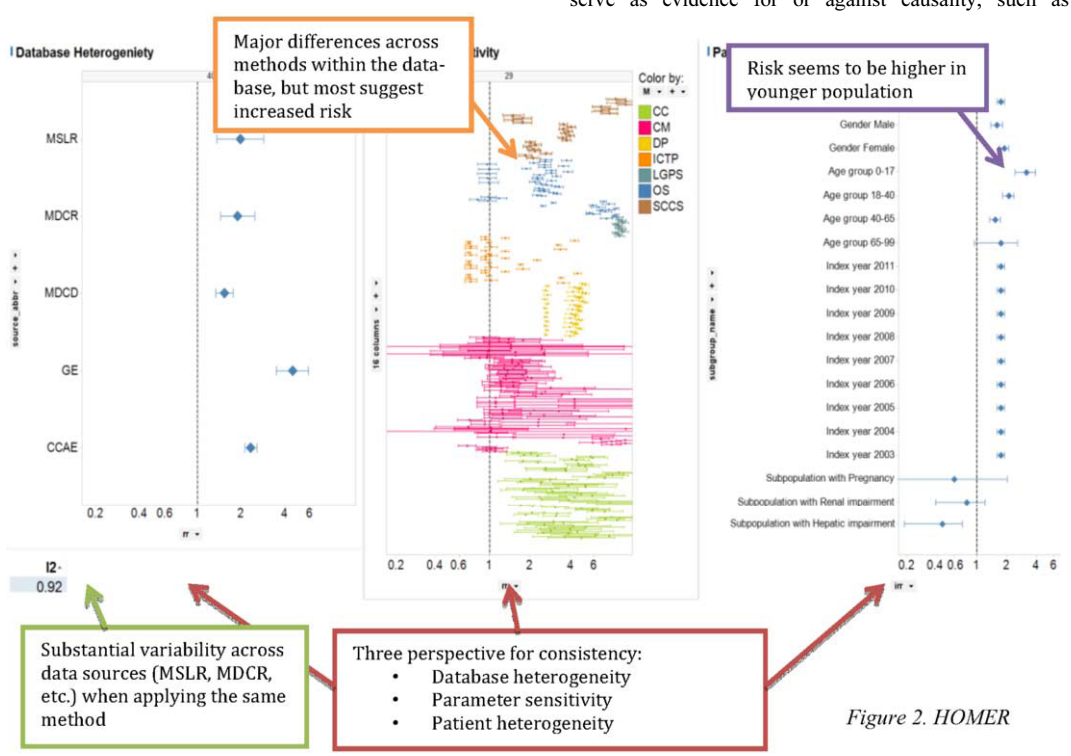


Figure 2. HOMER

consistency of findings. Consistency can be assessed across source datasets, analysis methods, or patient subpopulations. Figure 2 shows a prototype screen from HOMER, in which the relationship between a drug and adverse effect are explored across multiple dimensions.

In addition to creating user-friendly software, OHDSI also engages in the development of advanced analytic methods, including making Bayesian regression feasible over large data sets, handling sparse coding, and performing hierarchical association mining. Finally, OHDSI has been a leader in application of empircal calibration techniques to healthcare datasets to account for the biases inherent in these data sources as another contribution to the transparency and integrity of observational data research.

Research Network

The goal of the Research Network is to lower barrier to performing large-scale collaborative research using observational data and to generate high-quality evidence through peer review across study design, execution, and data analysis. Data owners make data available in return for algorithms and tools and the ability to post queries to the network. OHDSI has developed a process for the development and execution of research studies across the data network. This process includes proposal by a collaborator, review by the community, and promotion to active project. Initial proposals typically comprise a basic protocol including objective, rationale, target population, and initial source code. Once posted to the OHDSI Research Forum [8], the proposed project will be reviewed by other collaborators to determine interest level as well as to generate suggestions for modification to the design, target phenotypes, or analytic methods. Projects that generate interest from multiple sites and produce a complete protocol and cross-tested code are promoted to an Active Project. All community members are invited to run the analysis on their local dataset and return the results (de-identified, aggregate) centrally. Analyses across several sites rely on CDM standardization of various data domains to common terminologies, such as SNOMED for diagnoses, RxNorm for drug ingredients and LOINC for laboratory results. OHDSI's choice of standard terminologies does not limit who may participate because mappings and tools are supplied to translate from other terminologies. After a defined period, result submission is closed and the data are analyzed and presented back to the community. Based on these findings, publications and follow-up studies may result.

OMOP previously demonstrated that databases from different sources can give vastly different answers [9]. A blind aggregation of the results may increase variance due to heterogeneity instead of decreasing it due to sample size. Therefore, depending on the research question and the participating sites, the collected results may simply be reported without aggregation, summarized, or, if sufficient homogeneity can be demonstrated, aggregated.

Results

OHDSI held its first annual meeting at Columbia University Medical Center in New York City on 16 and 17 of October 2014. Fifty-eight participants reviewed the vision and goals giving rise to the creation of OHDSI and formed working groups to address the common data model, vocabulary, knowledge bases, estimation methods, phenotype generation, clinical characterization, and cohort definition. The outcomes of the meeting included the following:

1. Confirmation of the commitment of data custodians to participate in federated research studies;
2. Decision to open the database to queries from external researchers under a formal process; and
3. Work progress on each of the working group areas.

In preparation for the meeting, the OHDSI team surveyed current users of the OMOP data model. They found that 58 existing OMOP databases have collectively converted 682 million patient records to the OMOP CDM. This large number includes both patients and sources that are duplicated across databases and also includes databases that are not currently participating in the OHDSI Research Forum. Nevertheless, the total count demonstrates the feasibility of imposing a strong information model and executing a CDM conversion on a number of records that would represent a significant fraction of the world's population. In other words, the OHDSI vision is, in fact, feasible today. Furthermore, we estimate that the actual number of patients currently available in the OHDSI Research Forum is over 200 million.

OHDSI has just begun to distribute research queries. The first published OHDSI study used databases maintained by one site to carry out a medication-wide association study, assessing whether drugs with similar function or structure cause similar side effects [10].

Discussion and Conclusion

We envision a future where observational studies will inform clinical practice—by providing practice-based evidence—using the unprecedented amount of available patient data and the use of computerized systems to process the data [11]. Feinstein et al. initiated the idea of using data on 678 lung cancer patients as an electronic 'library of clinical experience' to obtain a personalized prognosis [12]. We believe that it now should be possible to provide not only prognosis information but also provide estimates on the comparative effectiveness as well as patient level assesment of treatment options.

We are not alone in this endeavor. For example, PCORnet [13] is designed to improve the national infrastructure for conducting clinical outcomes research. The network will enable a national capacity to conduct comparative effectiveness research efficiently and to learn from the health care experiences of millions of Americans.

OHDSI is implementing such a collaboration internationally by building on the OMOP experience for observational research. The existence of multiple very large databases that use the OMOP Common Data Model demonstrates that a worldwide conversion of clinical data from all patients is in fact feasible. Previous success in algorithm development, software distribution, and evidence generation points to potential success in the overall evidence-generating OHDSI project.

Acknowledgments

This work was funded in part by US National Science Foundation grant NSF IIS 1251151.

References

[1] OHDSI. http://ohdsi.org/ Accessed 2014-12-07.
[2] OHDSI Vision. http://ohdsi.org/who-we-are/mission-vision-values/ Accessed 2014-12-07.
[3] OHDSI Values. http://ohdsi.org/who-we-are/mission-vision-values/ Accessed 2014-12-07.
[4] Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety

surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60. Epub 2011 Oct 28.
[5] IMEDS. http://imeds.reaganudall.org/ Accessed 2014-12-07.
[6] ACHILLES Repository. http://github.com/ohdsi/achilles Accessed 2014-12-08.
[7] ACHILLES Demonstration. http://ohdsi.org/web/achilles/ Accessed 2014-12-08.
[8] OHDSI Forums http://forums.ohdsi.org Accessed 2014-12-08.
[9] Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage JM, Hartzema AG, Suchard MA, DuMouchel W, Berlin JA. Evaluating the impact of database heterogeneity on observational study results. Am J Epidemiol. 2013;178(4):645-651.
[10] Ryan PB, Madigan D, Stang PE, Schuemie MJ, Hripcsak G. Medication-wide association studies. CPT: Pharmacometrics & Systems Pharmacology 2013;2,e76;doi:10.1038/psp.2013.52.
[11] Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. Health Aff (Millwood), 2014;33(7):1229-1235.
[12]Feinstein AR, Rubinstein JF, Ramshaw WA. Estimating prognosis with the aid of a conversational-mode computer program. Ann Intern Med. 1972;76(6):911-921.
[13]PCORnet: the national patient-centered clinical research network. http://www.pcornet.org, Accessed 2014-12-08.

Address for correspondence

George Hripcsak, MD, MS
Profesor and Chair of Biomedical Informatics
Columbia University Medical Center
622 W 168th St, PH20
New York, NY 10032
hripcsak@columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-579

Identifying Repetitive Institutional Review Board Stipulations by Natural Language Processing and Network Analysis

Fabrício S. P. Kury, MD, James J. Cimino, MD

National Library of Medicine, US National Institutes of Health, Bethesda, Maryland, USA

Abstract

The corrections ("stipulations") to a proposed research study protocol produced by an institutional review board (IRB) can often be repetitive across many studies; however, there is no standard set of stipulations that could be used, for example, by researchers wishing to anticipate and correct problems in their research proposals prior to submitting to an IRB. The objective of the research was to *computationally identify the most repetitive types of stipulations generated in the course of IRB deliberations. The text of each stipulation was normalized using the natural language processing techniques. An undirected weighted network was constructed in which each stipulation was represented by a node, and each link, if present, had weight corresponding to the TF-IDF Cosine Similarity of the stipulations. Network analysis software was then used to identify clusters in the network representing similar stipulations. The final results were correlated with additional data to produce further insights about the IRB workflow. From a corpus of 18,582 stipulations we identified 31 types of repetitive stipulations. Those types accounted for 3,870 stipulations (20.8% of the corpus) produced for 697 (88.7%) of all protocols in 392 (also 88.7%) of all the CNS IRB meetings with stipulations entered in our data source. A notable peroportion of the corrections produced by the IRB can be considered highly repetitive. Our shareable method relied on a minimal manual analysis and provides an intuitive exploration with theoretically unbounded granularity. Finer granularity allowed for the insight that is anticipated to prevent the need for identifying the IRB panel expertise or any human supervision.*

Keywords:

Cluster Analysis; Natural Language Processing; Ethics Committees.

Background

For a majority of research proposal submissions the corrections to the text proposed by an Institutional Review Board (IRB), called "stipulations", are often stereotypical [1, 2]. Though these stipulations recur in a large number of protocols they can be categorized into a fewer common types. These repetitive problems can often be stylistic and editorial rather than regulatory [2], indicating that common problems, though undesirable, might be have an easy resolution. Automating the detection of stipulation-producing conditions in research proposals could streamline the review process. It could even be used to help researchers to correct the issues before submiting their proposals to the IRB. The existing work on identifying repetitive IRB stipulations was based on manual review [1] and thus burdensome to reproduce, limited in its breadth (total number of protocols and stipulations analysed), and low on granularity (how specific the results were). In this study we present a semi-automated and reproducible method for identifying repetitive stipulations, which allows intuitive exploration of its results, relies on minimal manual work, and does not use commercial software.

Introduction

This study sought to develop a method that identifies the most frequently occuring IRB stipulation types and is easily reproducible at other institutions using open access software. We used an approach to design a method for text clustering suitable to the stipulations' corpus.

A wealth of algorithms exist for text clustering [3]. However, they are either available as programming libraries for use in larger applications or as features of specialized commercial software that are rarely optimized for the characteristics inherent to very small texts – as are the individual IRB stipulations (which are often single sentences). Moreover, each algorithm has different tradeoffs and configurable parameters [3] that require advanced expertise. The available metrics for assessing the quality of the generated clusters is also an advanced topic. Effective visualization of the results is a task disconnected from the clustering itself.

The modeling of our text clustering task as a network cluster represented a way around the referred hindrances. The network analysis community is mature and provids a wealth of free software [4] with graphical user interfaces.Likewise, network clustering is a well-developed field of research with an array of free, and sometimes peer-reviewed, algorithms for use within the various network analysis suites. If a network could be built to truthfully represent the similarities of stipulations, its node clusters would represent the text clusters in graphical and user-friendly manner.

Method

Our methodology has five consecutive parts: A) Raw data retrieval and parsing; B) normalization of the text of the stipulation; C) construction of a network based on their similarities; D) identification of clusters in the network; and E) summarization the clusters for easier reporting. Parts A, B and C were performed by one Java program written for this study. Parts D and E were done manually using Cytoscape [7], a free network analysis software.

Raw data access and parsing

The raw data source for this study was the Combined Neurosciences (CNS) IRB database maintained by the NIH's Protocol Tracking and Monitoring System (PTMS), which was launched in 2005 but contains data dating back to 1999. As of December 15th, 2014, the PTMS database contained 786 protocols for the CNS that had been reviewed in 442 meetings of the three CNS IRB panels and with stipulations entered in the system.

In PTMS, all stipulations produced for a given protocol in a meeting are stored as a continuous narrative text. The structured text of the stipulations (each stipulation is written on a new line), enabled the use of regular expressions to parse and separate them individually. The proportion of incorrectly parsed stipulations was auite low because of their dissimilarity to all others, those would tend to not interfere with the rest of the results. The availability of separated stipulations was therefore assumed for this study.

Normalization of the stipulation text

To mitigate the inherent variability of the natural language, each stipulation was normalized into a set of non-repeating terms ("a bag of terms"). The normalization process focused on preserving the conceptual content of the stipulation subject while systematizing the ways in which the content could be verbalized by removing minor details. The repeating terms in small text strings of varying structures were not likely to represent different core concepts and therefore were considered only once.

The normalization part made principal use of the freely available Stanford Core NLP (SCNLP) library for Java [5]. In a six-step process, the normalization 1) removed personal names and dates identified by SCNLP's Named Entity Recognizer; 2) converted terms to their preferred synonyms, using a list of 9654 mappings from a previous unpublished work of our group (Table 1); 3) converted words to their lemmas as per the SCNLP's 'lemmatizer'; 4) removed 475 stop words that likely have a grammatical function; 5) applied the Porter Stemmer [6] to all terms; and 6) used the "\W" regular expression in Java to remove punctuation, tabs, special characters, etc. Table 2 provides examples of stipulations, their normalized versions, and calculated similarities.

Construction of an undirected weighted network based on the similarities among the stipulations

A network node, with a unique ID, was created for each stipulation . The nodes were numbered cardinally (1, 2, 3,…) wothout regard to the order of their creation. Each node also contained a "Stipulation property", with the original unmodified stipulation text that it represents, to allow its reading from the network analysis software. The TF-IDF Cosine Similarity algorithm [3] was then used to calculate the similarity (a number from 0 to 1) between each normalized stipulation to each one of the others. An undirected link (or "edge") was created between two similar nodes with a weight property that is a mathematical inverse of the calculated similarity ($weight = \frac{1}{similarity}$). Links with similarity lower than a predefined minimum similarity threshold (MST) were not created. Though the network clustering algorithm used in part C takes the link weight into consideration when identifying the clusters, links with similarity lower than a predefined minimum similarity threshold (MST) were not created to expedite the processing. After iterative testing the MST was set at 0.15, which, while imperceptible to human reader, significantly reduced the total number of links in the network.

Table 1 – Examples of preferred synonym mappings

Term	Preferred synonym
GUIDELINE	RULE
GUIDELINES	RULE
HEADACHE	CEPHALALGIA
VII	7
VIII	8
TOOTHS	TEETH
TOOTH	TEETH
MALIGNANCY	NEOPLASM
NEOPLASTIC	NEOPLASM
MALIGNANT	NEOPLASM

Identication of clusters in the network

The similarity-interlinked stipulation network is produced as a GraphML file by the Java program, which is then loaded into the free Cytoscape [7] software for the final task of identifying the clusters. In Cytoscape we opted to use the free plugin ClusterONE, which provides the official implementation of a peer-reviewed algorithm [8] to identify overlapping clusters in weighted networks.

We configured ClusterONE parameters so as to deem a stipulation repetitive if it appeared in more than 10% of all the 782 protocols, making our minimum cluster size to be 79. We left all other parameters at the predefined levels. We would only accept clusters with a p-value ≤ 0.05, as calculated by the Mann-Whitney U test provided by ClusterONE. As preconfigured in ClusterONE, the clusters were allowed to overlap each other.

Summarization of the clusters

As the task of manually reviewing the entire content of each cluster would be prohibitive, undermining the reason of using computational methods in first place, the summarization of each cluster also had to be carried out computationally. From the network analysis perspective this is a simple task. Since the links are weighted according to the corresponding stipulation similarity, it follows that the nodes with the highest weighted closeness centrality (WCC, Figure 2) correspond to the stipulations with closest similarity to all other stipulations in the cluster. Therefore, each cluster identified in part D was cloned as a separate network; we then used the free Cytoscape plugin CytoNCA [9] to calculate WCCs. If two clusters happened to share the same highest WCC node, they were merged using the union merge command in Cytoscape, which recognized identical nodes by their identical ID numbers preventing their duplication. From each of the final clusters, the node with the highest WCC was reported in Table 3.

Table 2 – Examples stipulations and normalization

Stipulation	After normalization	Similarity to above[1]
Check/specify the benefit level.	specifi benefit level	-
Provide a plan to increase recruitment of minority participants.	provid plan increas recruit minor particip	0.0
Clarify the accrual ceiling. Provide a rationale for the request for the increase ceiling.	clarifi accrual ceil provid rational request increas	0.091020 *(rounded)*
Clarify the new accrual ceiling.	clarifi accrual ceil	0.273721 *(rounded)*
State that Christopher Humphrey[2] will not obtain consent.	obtain consent	0.0
State whether Dr. Johnson[2] will obtain consent.	obtain consent	1.0
State who obtains consent.	obtain consent	1.0

[1] TF-IDF Cosine Similarity with IDF weights as computed from our corpus of 18,582 stipulations. [2] Fictitious names.

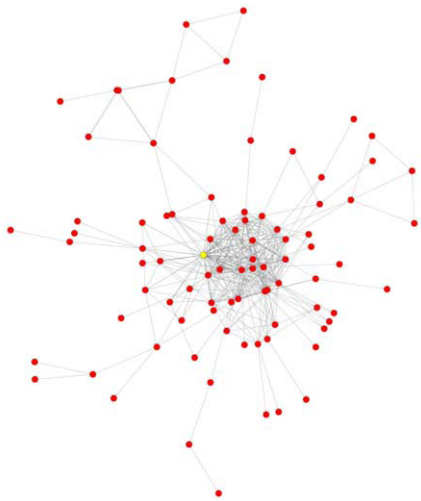


Figure 1 – Group 31, example of a sparse cluster, demonstrating high variability in the wording of the stipulations included. The node with highest WCC is colored in yellow: "Make study population numbers consistent with CR memo and protocol."

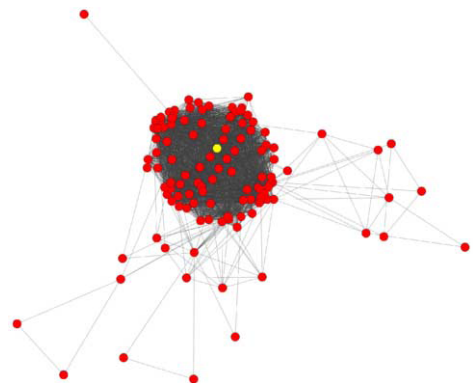


Figure 2 – Group 7, example of a dense cluster, demonstrating low variability in the wording of the stipulations included. Node with highest WCC (in yellow): "Correct typographical errors."

Table 3 – Types (Groups) of repetitive stipulations identified in our studied corpus

	Node with highest weighted closeness centrality	Occurrences[1]	Cluster quality[2]
1	Revise language to reflect the Board's determination of unanticipated problem and adverse event reporting requirements for this protocol. Language will be provided by IRB staff.	196	0.872
2	Assure that the new investigator has completed NIMH HSPU consent training before obtaining consent.	82	0.608
3	Provide FDA documentation for the IND when it becomes available.	236	0.602[3]
4	Provide a rationale for the upper age limit.	100	0.59
5	Change the protocol risk level to more than minimal risk.	159	0.586
6	Provide a plan to improve recruitment of minority subjects.	159	0.556
7	Correct typographical errors.	121	0.546
8	Specify the total amount of blood drawn for this study and what the blood will be used for.	118	0.545
9	Clarify and make consistent the number of patients and healthy volunteers.	120	0.544
10	Change the classification of these events to a serious unanticipated problem and a serious deviation, per the IRB's determination.	91	0.54
11	Clarify where samples and data will be stored.	144	0.538
12	Clarify the compensation sections.	145	0.53
13	Update standard language for MRI procedure and risks.	174	0.511
14	Make the accrual numbers consistent across all documents.	185	0.505[3]
15	Submit FDA annual report.	103	0.503
16	Provide the rationale for excluding minors.	136	0.479
17	Change "your can" to "you can".	87	0.472
18	Go over the statistical review with Dr. Miller[4] and clarify the analysis plan.	92	0.471
19	Add inclusion and exclusion criteria.	239	0.469
20	Correct the contact information for Dr. Miller[4].	99	0.448
21	Add information about pregnancy testing in minor females.	198	0.44
22	Change "will not" to "may not" provide direct benefits.	110	0.439
23	Adverse Event Reporting, clarify which expected drug effects will not be reported as adverse events.	178	0.436[3]
24	Update the protocol and consents to the current CNS templates by the next CR.	93	0.432
25	Specify who will be responsible for individual subject safety monitoring.	84	0.413
26	Update the list of investigators.	94	0.408
27	Specify which investigators can obtain consent for this study, and provide their qualifications.	114	0.396
28	Work with IRB staff on simplifying language.	221	0.389[3]
29	Specify the risk and benefit levels.	118	0.365
30	Make sure information about genetic testing is consistent with the protocol.	104	0.361
31	Make study population numbers consistent with CR memo and protocol.	88	0.347

[1] Numbers are approximate and Groups can overlap each other. [2] As measured by ClusterONE [8]; [3] Value of the biggest cluster before merging. 4: Names are fictitious.

Results

We parsed 18,582 stipulations from 786 protocols and 442 IRB meetings (examples in Table 2). A network with 124,219 links was generated for the minimum TF-IDF Cosine Similarity seting of 0.15. When seen in its entirety the network was cluttered and opeque, with no visual clusters. In the first execution, ClusterONE identified 36 clusters larger than 79 nodes, 29 of them with a p-value below 0.001, six below 0.05, and one above 0.05 (0.083) which was discarded. Four cluster pairs were found to share their node with highest WCC and were therefore merged (clusters 3, 14, 23 and 28 in Table 3). The *cluster quality*, as measured by ClusterONE (with the stronger connections being inside the cluster), ranged from 0.347 to 0.872. Altogether, the 31 final clusters (after merging 4 of them) contained 3,870 stipulations (20.8% of the corpus) produced for 697 (88.7%) of all protocols in 392 (also 88.7%) of all the CNS IRB meetings with stipulations entered in the PTMS.

Discussion

Modeling the problem as a network permitted instant visual assessment and exploration of the results. The visualization of the clusters in link (edge)-weighted layouts, which position each node according to its links and their weights, provides intuitive insight about the collective characteristics of the stipulatons and the existance of the cluster subgroups. Without a visual resource it is cumbersome to assess the variation among the stipulations.

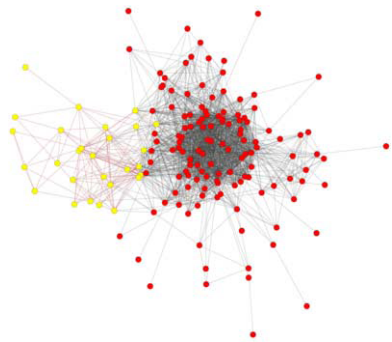
Defining 79 repetitions (over 10% of all protocols) to deem a stipulation stereotyped was a conservative approach. Reducing that parameter to 40 resulted in 96 initial clusters with p-value ≤ 0.05 (out of a total 105), however, the smaller clusters pose unnecessary separate representations of the same core term as understood by a human reader. This minimum cluster size parameter, as well as the others provided by ClusterONE, can be used for deeper investigation: for example, to subdevide an initial large cluster into smaller ones using lower minimum cluster sizes (Figure 3).

Limitations

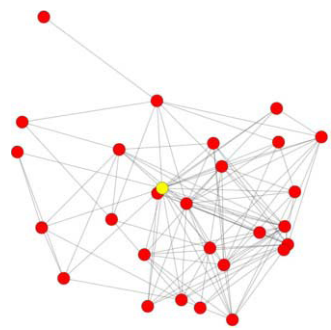
The text in Table 3 may be taken as a representation of the stipulation types, however, there may be other variations. For example, Group 5 contained stipulations stating the exact opposite of the one cited in the Table, and Group 15 included the same stipulation for other specific cohorts such as HIV-positive persons, cognitively impaired persons, persons who cannot give consent, and so on. Figure 3 provides an example of exploring the internal variability of a group. The granularity of the final results is theoretically indefinite with visual navigation, however, the realization of the presence of sub-clusters is not as easy due to their proximity. Nevertheless, the proximity of similar stipulations facilitates identification of sub-clusters by trial-and-error.

The boundaries of the clusters and sub-clusters are often not precise (compareGroups 16 with 19, and 22 with 29 in Table 3). The cluster-size, therefore, is an approximation. These imperfections were not investigated in this study beyond the cluster quality measurements provided by our chosen plugin ClusterONE.

The network analysis is only as meaningful as the links between the nodes that represent the conceptual similarities between the stipulations. While the TF-IDF Cosine Similarity algorithm provides different weights for the words according to the frequency of their occurrance the stipulation text



Group 5 with sub-cluster (in yellow) as identified by running ClusterONE again with minimum cluster size of 25.



Node with highest WCC in sub-cluster (in yellow): "Specify the risk category for minor participants."

Figure 3 – Sub-clustering Group 5.

variability was essentially lexical. More elaborate methods for quantifying the semantic similarity between pieces of text do exist [10], however, their computational efficiency is not practical to cross-compare thousands of text strings (in this study, the production of the GraphML file took the Java program about 45 minutes on a dual-core 3.33 GHz/8 GB of RAM desktop computer). Finally, the use of a preferred synonym dictionary for reducing the lexical variablity – an extremely fast implementation approach available at our laboratory – can introduce a layer of semantic matching between the words; but a more robust and appropriate method would be to use a lexical database, such as WordNet, to get a normalized common denominator calculating the word-level similarity [11].

Clinical research implications

The high incidence of stereotyped stipulations in a majority of the protocols indicates insufficiency of the review process before submition to an IRB, and the repetition types suggest that their management does not require IRB-level expertise. Our method relies on a minimal manual effort and is expected to be reusable at other institutions. Our group will be glad to share our written Java program upon request.

Clinical research informatics implications

Several of the most recurring stipulation stereotypes can be computationally identified from the research protocol text. For example, Groups 7 and 17 in Table 3 could be resolved using spellcheck. Group 13, which demands the use of standard language, could be resolved by algorithms that identify the missing expressions. Group 28 could be addressed by a software that gauges the level of vocabulary and raises an alert

if it exceeds a predefined. Others (8, 9, 11, 14, 19, 27, 29) could be addressed by more elaborate protocol authoring tools involving structured data entry of the elements causing the stipulations.

Conclusion

We have successfully identified the most repetitive IRB stipulation types at lexical level with a hybrid approach that consists of computational algorithm using a free network analysis s/w and some manual effort. Our method allows indefinite breadth of investigation and granularity (detailing) amenable to a lay user thorough intuitive visual exploration. The knowledge, and subsequent resolution of, the most common problems can help authors to avoid delays in their protocol approvals. The IRBs can improve their own efficiency by developing strategies to identify the repetitive problems in protocols before initiating a review meeting. We believe that some of the most repetitive problems we have identified warrant further interest on the usefulness of computational analysis of the research protocols.

Acknowledgements

This work was supported by intramural research funds from the NIH Clinical Center and the National Library of Medicine. The authors express their gratitude for the PTMS database access kindly authorized by Dr. Barbara Karp, chair of the Combined Neurosciences IRB of the NIH.

References

[1] Sansone RA, McDonald S, Hanley P, Sellbom M , Gaither GA. The stipulations of one institutional review board: a five year review. J Med Ethics 2004;30:308–310.

[2] Abramovici A, Salazar A, Edvalson T, et al. Review of multicenter studies by multiple institutional review boards: characteristics and outcomes for perinatal studies implemented by a multicenter network. Am J Obstet Gynecol 2014;211:xx-xx.

[3] Aggarwal CC, Zhai CX. A Survey of Text Clustering Algorithms. Mining Text Data 2012, pp 77-128.

[4] http://en.wikipedia.org/wiki/Social_network_analysis_software. Accessed December 18th, 2014.

[5] Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, and McClosky D. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

[6] Porter MF. An algorithm for suffix stripping. Program, 14, no. 3, pp 130-137, July 1980.

[7] http://cytoscape.org/. Acessed December 18th, 2014.

[8] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. 2012. Nature Methods, vol.9, no.5, pp 471-475.

[9] Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of biological networks. BioSystems, 2014, DOI: 10.1016/j.biosystems.2014.11.005

[10] Rus V., Lintean M., Banjade R., Niraula N., and Stefanescu D. 2013. SEMILAR: The Semantic Similarity Toolkit. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, August 4-9, 2013.

[11] http://wn-similarity.sourceforge.net/. Acessed December 18th, 2014.

Address for correspondence

Fabrício S. P. Kury, M.D.

Laboratory for Informatics Development

US National Institutes of Health Clinical Center

Office: +1 301-435-3183, e-mail: ***fabricio.kury@nih.gov***

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-584

Patient-Centered Outcomes Research in Practice: The CAPriCORN Infrastructure

Anthony Solomonides[a], Satyender Goel[b], Denise Hynes[c,l], Jonathan C Silverstein[a], Bala Hota[d1], William Trick[e], Francisco Angulo[e], Ron Price[f], Eugene Sadhu[c], Susan Zelisko[f], James Fischer[c], Brian Furner[g2], Andrew Hamilton[h], Jasmin Phua[i], Wendy Brown[j], Samuel F Hohmann[k,d2], David Meltzer[g1], Elizabeth Tarlov[c,l], Frances M Weaver[f,l], Helen Zhang[e], Thomas Concannon[m], Abel Kho[b]

[a] Center for Biomedical Research Informatics, NorthShore University HealthSystem
[b] Feinberg School of Medicine, Northwestern University
[c] University of Illinois, Chicago
[d]Rush University Medical Center—[1]Department of Medicine, —[2]Dept of Health Systems Management
[e] Cook County Health and Hospital Systems
[f] Loyola University Health System
[g] University of Chicago—[1]Medicine, —[2]Center for Research Informatics
[h] Alliance of Chicago Community Health Services
[i] Medical Research Analytics and Informatics Alliance
[j] VA Jesse Brown Hospital
[k] Universities Healthsystem Consortium
[l] VA Edward Hines Hospital
[m] RAND Corporation.

Abstract

CAPriCORN, the Chicago Area Patient Centered Outcomes Research Network, is one of the eleven PCORI-funded Clinical Data Research Networks. A collaboration of six academic medical centers, a Chicago public hospital, two VA hospitals and a network of federally qualified health centers, CAPriCORN addresses the needs of a diverse community and overlapping populations. To capture complete medical records without compromising patient privacy and confidentiality, the network created policies and mechanisms for patient consultation, central IRB approval, de-identification, de-duplication, and integration of patient data by study cohort, randomization and sampling, re-identification for consent by providers and patients, and communication with patients to elicit patient-reported outcomes through validated instruments. The paper describes these policies and mechanisms and discusses two case studies to prove the feasibility and effectiveness of the network.

Keywords:

Patient-Centered Outcomes Research; Comparative Effectiveness Research; Electronic Health Records; Data Collection; Data Linkage; Aggregation; Data Sets; Deidentification;Re-identification; Consent.

Introduction

PCOR, CER and PCORnet

The Patient-Centered Outcomes Research Institute (PCORI) was established following the US Patient Protection and Affordable Care Act in 2010. PCORI's mission is to advance and support Patient-Centered Outcomes Research (PCOR), which "helps people and their caregivers communicate and make informed healthcare decisions, allowing their voices to be heard in assessing the value of healthcare options." [1]

In particular, PCOR:

- Encompasses comparative effectiveness research (CER) on interventions to inform decision making.
- Addresses individuals' (especially patients' and care-givers') preferences and autonomy.
- Studies a diversity of settings and populations.
- Seeks to balance stakeholders' concerns, including burden to individuals and availability of resources.

One principal action of PCORI is to support 11 Clinical Data Research Networks (CDRN) and 18 Patient-Powered Research Networks (PPRN). Both kinds of research networks are seen as infrastructure-building projects, with specific structural, process and outcome goals to prove the feasibility and usefulness of the networks. CDRNs focus on major academic medical centers: apart from demonstration of viable infrastructures, CDRNs demonstrate their value by conducting research in a number of specific conditions. Each network nominates the conditions on which it will work. However, longer term sustainability for the infrastructure can only be achieved through success in early studies, and proving to the research community that the network represents a valuable resource that is worth both exploiting and supporting through further funded studies and grant proposals. PPRNs focus on specific conditions that are of concern to patients, care providers, and patient advocacy organizations. Many networks have formed around existing formal or informal networks of support and advocacy groups.

Overarching the CDRNs and PPRNs, PCORI established a supra-network, PCORnet, that acts as a collaboration venue, clearing house, and policy-development body. Best conceived of as a network of networks, PCORnet ensures that the infrastructures created by the different CDRNs and PPRNs will remain interoperable and responsive both to researchers' needs and to the expectations of patients, care providers and advocates.

CAPriCORN

One of the CDRNs, CAPriCORN, represents an alliance of Chicago institutions collaborating in recognition of the need for pre-competitive comparative effectiveness research (CER) in their highly diverse community—diverse both in the type of

institutions involved and, importantly, in the populations they serve. CAPriCORN is not typical of CDRNs, although it shares many characteristics. Some of its unique features provide a model for collaboration in environments where, for example, patient populations at different institutions overlap, where nevertheless a full picture of each patient's health record is necessary for meaningful research results.

Data for sharing within CAPriCORN—and in the wider community at a later stage—will be in a HIPAA-compliant, de-identified format. Two working groups (WG), Informatics WG and Ethics and Regulatory WG, devised a federated data architecture, a data model with appropriate standards, and a designed data flow engineered to ensure that no protected health information (PHI) is released other than under strictly controlled conditions and, at the same time, maintaining the research value of the data that is released. De-identified data will be released on a study-by-study basis. A statistically benchmarked process is used to generate a pseudonymous identity for each patient in such a way that distributed patients' records across different providers in the network can be matched and integrated. The records are not brought together into a single central database, but are instead put in a virtual repository – by allowing distributed queries across the different systems through the validated mechanism of PopMedNet [4, 5]. Consent will be sought when access to PHI, or directly to the patient for patient-reported outcomes, is necessary.

Methods

Population

CAPriCORN comprises a network of six academic medical centers (University of Chicago, University of Illinois, Chicago, Loyola University, NorthShore University HealthSystem, Northwestern University and Rush University Health), the Alliance of Chicago's Federally Qualified Health Centers, a major public hospital, Cook County Hospital, and two Veterans Affairs hospitals, VA Edward Hines and VA Jesse Brown. Geographically, these institutions serve the greater Chicago metropolitan area and are available to a total population of approximately 9.5 million. (In addition to these "data-providing" institutions, 22 other organizations contribute research, patient advocacy, and infrastructure services to CAPriCORN. Their role is described below.)

CAPriCORN institutions together held 2,860,000 covered lives in electronic health records. A preliminary analysis of seven of the ten institutions indicated 6,923,111 patients, of whom 1,465,285 were registered with a primary care provider; however, after de-duplication, the numbers were 5,741,268 and 1,242,380 unique patients respectively. Thus some 20.6% of patients are associated with more than one institution, and even among the primary populations, there are 18% of patients with more than one PCP registration. This appears to be symptomatic of deprivation in the inner city, where economic necessity requires individuals to move opportunistically from provider to provider.

The racial breakdown of the primary population is 47.5% Caucasian, 27.9% African American and 14.9 Hispanic, with just over 9% in other categories. Of this population, 59.3% are female, 40.7% male. The mean age is 50 with a standard deviation of 17.9.

De-identification and De-duplication

While fragmented care may be suboptimal, research on comparative effectiveness of treatments requires as accurate and as complete a record of each patient's health status and episodes of illness as can be reconstructed, if meaningful and valid results are to be achieved. With multiple records for up to 20% of patients, de-duplication is strongly indicated. The means of achieving this lie in a method of de-identification.

In the US, there is currently little prospect of a single unique patient identification code. Where health information exchanges have been instituted, it is necessary to implement an "enterprise master patient index" (EMPI), but even these are rare because of a number of concerns, principally privacy and security, and economics and sustainability. Nevertheless, prior experience was sufficiently encouraging to suggest that a specific design and implementation in the Chicago area would be worthwhile. This prior knowledge and experience provided a fundamental cornerstone for the CAPriCORN network.

The de-identification algorithm comes from Kho *et al* [2, 3]. The algorithm uses a set of strictly personal identifiers, i.e., nothing that may be institution-specific, to generate up to 17 different combination strings and uses a statistically selected subset of these to construct a "hash-ID." The hashing algorithm is not reversible, but its high specificity allows patients who have multiple records to be discovered, albeit anonymously.

Organizational Design

CAPriCORN is led by a Principal Investigator at the Chicago Community Trust, an organization focused on civic leadership and philanthropy. A Steering Committee is the decision-making body, whose composition was designed around the natural concerns of a network to conduct and facilitate patient-centered outcomes and comparative effectiveness research across a number of healthcare institutions. The Steering Committee also reflects the underlying architectural design of the infrastructure and the projected governance and regulatory framework of that infrastructure.

Clinical Data Research Networks are intended to be open to external collaboration, explicitly designed to be open to patient concerns, and subject to all the normal ethical and regulatory processes that apply to human subjects and social science research. These are, respectively, reflected in the network's External Researcher Committee, Patient and Clinician Advisory Committee, and Chicago Area Institutional Review Board (CHAIRb). All these committees define processes and workflows for patient and carer consultation, the triage of internal and external research proposals, the handling of data requests, the release of data, and the consenting process prior to any re-identification of and contact with patients.

Critical to the infrastructural design are two "honest broker" roles in the network. Other than in very specific, precisely defined circumstances involving only consented patients, these organizations hold no protected personal health information (PHI) but handle the "de-identifiers", principally the hash-IDs for de-duplication, and subsequent to the definition of specific condition cohorts, a second level of pseudonymization, the cluster-IDs, which are randomly generated "per study, per hash-ID" thus avoiding any unintended crosstalk between independent studies.

The principles, explicit and implicit, that guided this design are:

- All studies, including those submitted as "proof of principle" for the network, along with new and external proposals, will be subject to triage by the Patient and Clinician Advisory and External Researcher committees, then subject to review by CHAIRb, with the ultimate decision resting with the Steering Committee.

- All PHI will be held at institutions, benefiting from all the protections (firewalls, authorizations, etc.) that each applies to its own patient data.
- The data collected will be strictly non-PHI and minimal with respect to any cohort identification needs (all that is needed, but no more).
- Identifiers will be hashed into pseudonymous "hash-IDs" for the purpose of de-duplication. Honest Broker 1 (HB1) will provide institutions with a unique "hash seed" that each will use to de-identify its own patients through hashing.
- The second honest broker, HB2, will use the hash-IDs provided by institutions to identify "duplication" and determine the set of institutions to which each patient corresponds. HB2 then generates a random identifier, the cluster-ID, for each unique patient in the given cohort. At this point, if considered necessary, the institutions themselves may be pseudonymized. (No PHI will flow to HB2.)
- Patients' records may only be linked through the hash-ID. Cohort identification for specific studies and non-PHI data requests from sites for constructing aggregate records may be conducted only by means of a distributed query mechanism (currently, PopMedNet [4, 5]) which allows inspection and vetting of queries prior to execution and results from queries to be examined prior to release.
- All studies that require access to PHI must identify a co-investigator at each site.
- Provider consent to approach patients to consent for particular studies will be requested, and subsequent patient consent will be sought, according to institutional rules and norms.
- Randomization of patients for consent will be done anonymously both in respect to patients and institutions.

As noted above, these principles are visible in the organizational structure of the network, but they are also evident in the architectural design of the infrastructure.

Network Architecture

The architecture and processes represented by the various flows in this diagram are detailed in Figure 1.

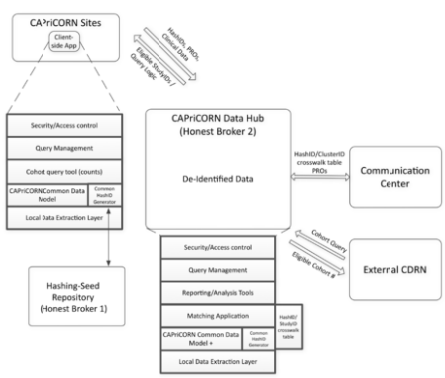


Figure 1 – A schematic diagram of the network displaying the two "honest broker" roles, the institutional repositories and the central "data hub" which hosts the matching and distributed query services.

CAPriCORN developed a data model and data standards, together with "extract-transform-load" processes for its institutional data marts. The data model is effectively based on a star schema with the concept *Encounter* at its center, so that data can be understood at a transactional level. A data dictionary was adopted showing domains and variables within them (apart from patient demographics, radiating out from encounters are diagnoses, medications, procedures, vital signs, laboratory results, and some additional local variables). Standards and terminologies indicate values in each category. The degree of privacy restriction for each variable (within-institution, within-CAPriCORN, within-PCORnet) is also indicated.

Each institution established a data mart (or other local database) which, notwithstanding the differences in platforms, precisely matches the CAPriCORN data model. Thus, although local adaptations of SQL queries will be necessary, the essential logic of queries submitted to the "data hub", i.e., the distributed query service, will remain unaltered, as required by PCORnet for its greater vision of seamless patient-centered, comparative effectiveness research.

A Communication Center is also being established to facilitate the process of re-identification of patients for provider consent to approach patients and for patient consent to participate in survey research (patient-reported outcomes, or PROs) and intervention studies. Each institution's processes are respected, and no pre-consent PHI flows through the center.

Process Description

1. HB1 hosts a stand-alone, generic hashing-seed generator application; it generates a SEED and passes it automatically to all participating institutions.
2. Each INSTITUTION uses the SEED and a set identifiers to generate a set of multiple *hashes* for each patient on record:

   [SSN, FirstName, LastName, DoB, Gender] ⊗ SEED
   → { hashes }

   from which a unique hash-ID is generated and cross-linked to the patient's MRN for internal identification.

This is per patient; […] signifies a vector of personal data.

Hash-IDs can be used within each INSTITUTION locally, if desired.

3. For each STUDY, every INSTITUTION runs the appropriate phenotyping algorithm to select its subpopulation of all unique patients who satisfy the cohort criteria. The hash-IDs along with all the hashes are returned to HB2.
4. For each study, HB2 collects all hashed data and de-duplicates, storing the result in a vector as follows:

   { (institutionID=1) : hash-ID$_1$ }
   $\Diamond_{\text{hash-ID}}$ … $\Diamond_{\text{hash-ID}}$
   { (institutionID=10) : hash-ID$_{10}$ }
   → hash-ID : institutionVector

where $\Diamond_{\text{hash-ID}}$ represent the join on hash-ID. The patient's hash-ID and institutionVector now appear thus:

	Institutions									
Disease D	AL	CC	UC	UI	LU	NS	NU	RU	VH	VJ

hash-ID										
xyz123	0	0	0	1	0	0	0	1	0	0

The patient whose hash-ID is "xyz123" was identified as having disease D and having partial records at UI and RU. We note that

(i) the hash-ID is in reality a more complex object (cf. [2]);

(ii) this may not be the complete record for this patient.

5. The five collections { hash-ID }, one for each study, are returned to all the institutions for cohort verification.

This is necessary, because, for example, a patient with an anemia record at one hospital (RU) may turn out to have a record at another hospital (UI) that does not mention anemia. Nevertheless, a complete record for that patient must include the partial records from both institutions.

6. Each institution checks the lists against its reference hash-ID list and so completes each patient's record if necessary.

For the sake of illustration, suppose now that we have found the patient above has also been seen at yet another hospital (CC) for an unrelated condition. The vector now becomes:

Disease D	AL	CC	UC	UI	LU	NS	NU	RU	VH	VJ

hash-ID										
xyz123	0	1	0	1	0	0	0	1	0	0

We can now confidently compile a complete record of the patient.

7. At this point, HB2, as an honest broker, must do two more de-identification steps:

   a. disguise the institutions

   b. replace hash-IDs with non-derived ids for the patients; these are the cluster-IDs.

For the first step, HB2 randomly assigns pseudonyms to the institutions, say:

AL	CC	UC	UI	LU	NS	NU	RU	VH	VJ
ff	*dd*	*aa*	*jj*	*bb*	*ii*	*cc*	*ee*	*hh*	*gg*

and these are then indexed as:

aa	*bb*	*cc*	*dd*	*ee*	*ff*	*gg*	*hh*	*ii*	*jj*
UC	LU	NU	CC	RU	AL	VJ	VH	NS	UI

The example patient now appears as:

Disease D	*aa*	*bb*	*cc*	*dd*	*ee*	*ff*	*gg*	*hh*	*ii*	*jj*

hash-ID										
xyz123	0	0	0	1	1	0	0	0	0	1

   c. The hash-IDs for each study cohort can now be replaced with unique cluster-IDs.

Our example patient now appears as:

Disease D	*aa*	*bb*	*cc*	*dd*	*ee*	*ff*	*gg*	*hh*	*ii*	*jj*

cluster-ID										
D-900093	0	0	0	1	1	0	0	0	0	1

Now, only possession of the table converting hash-IDs to cluster-IDs can enable anyone to re-identify the patient.

Distributed Queries

With cohort cluster-IDs collected, HB2 routes data requests through the distributed query service to the institutional data marts (IDMs). Locally, each institution will determine if the proposed query against its IDM is acceptable, allow the query to execute, and even then scrutinize the results before releasing them. Both in sending the requests and as results are received, HB2 can match cluster-IDs to hash-IDs, so that even a clinician researcher working on a project in their own specialty may be able to view expanded records of their own patients without recognizing them as their own. This provides a very high standard of de-identification.

Re-identification

Once particular studies based on entire cohorts are launched, re-identification of subsets of patients will most likely be necessary. Having received approval both from the Steering Committee (with advice from PCAC and ERC) and permission to proceed from CHAIRb, a researcher may request the Communication Center to randomly select a possibly weighted sample from across institutional or other populations for re-identification. The researcher will also be able to submit, through HB2, a data request for controls. Subject to CHAIRb's approval, institutional processes can be employed to gain provider consent and from there patient consent to participate in a study. Given the cluster-IDs of the patients in the study group, the Communication Center can alert institutions to the hash-IDs of patients to be approached for re-identification. In some cases, the Communication Center will also provide institutions with the means to collect patient-reported outcomes.

In the case of patients attending multiple institutions, which institution (or more precisely, which provider) should consent the patient for an identified study may be complex. A variety of algorithmic approaches is possible, including some that may work well but are computationally expensive. This may take the form of querying the system for the number of encounters at each institution in the last year (complex, but likely to reflect the patient's expectation) or it may suffice to look where the patient is registered for primary care (inexpensive, but may be irrelevant). The present ruling of CHAIRb only constrains the approach to be through a provider who is actually involved in the patient's care.

Results

Approximately at the halfway point in the project, achievements across a number of fronts include:

- Establishment of a sound governance structure, including a common central IRB, with data use and business associate agreements in place.
- Establishment and launch of a Patient and Clinician Advisory Committee with a clear role in the review, triage, and approval of new research proposals and a comprehensive manual for its operations.
- Approved design for the technological infrastructure, including a data model designed for ease of distributed query as well as with model evolution in mind.
- Approved processes and workflows now increasingly described and approved in protocols.

- Preliminary tests of the de-identification process and the distributed query machinery.
- Preliminary phenotyping in all five study cohorts proposed at project submission (see below). Preparatory phenotyping for a number of other studies, including incidental findings in osteoporosis, the national aspirin trial, bariatric surgery, antibiotics and childhood weight, bisphosphonates, and others.
- The de-identification and de-duplication processes in CAPriCORN are increasingly being reviewed as a model to be replicated across other CDRNs.

The internal organization of the network lends itself well to establishing CAPriCORN as a corporate entity; this would no doubt present new challenges, but is under consideration.

Discussion

The data model deployed at institutions to construct a data mart. Based on model variables, five phenotyping algorithms were devised and tested at multiple sites to identify overweight and obese patients (as required of all CDRNs); ambulatory patients suffering from asthma and in-patients with anemia (the two common disease cohorts); and patients with recurrent *Clostridium difficile* infection (RCDI) and sickle-cell disease sufferers (the two rare conditions).

In preparation for all these studies (and other anticipated future studies, including the PCORnet-inspired Aspirin trial and various collaborations with other CDRNs and PPRNs) the central IRB, CHAIRb, reviewed a Master Protocol which serves as a prefix to all specific study protocols.

Extract-Transform-Load (ETL) processes were undertaken against a number of different proprietory EHR systems. Some of these were shared publicly (e.g., through an EHR vendor's community sharing portal, thus conforming with requirements of commercial confidentiality). ETL logic was shared among all data-contributing sites to ensure compatibility.

The CAPriCORN data model is a superset of the PCORnet common data model against which external requests will be formulated. This model produces a straightforward mapping of data and requests from PCORnet to CAPriCORN. Additional data models influence the central PCORnet design, such as (Mini-)Sentinel, OMOP, i2b2 and others, and studied with a view to establishing correspondences should collaboration make a translation between CAPriCORN and another data model desirable.

Among the proposed cohort studies, the case of RCDI provides a convenient example of a hard test-case for the infrastructure. The study has not yet been completed, but based on data stored according to the data model and addressing queries to pre-existing institutional data warehouses rather than the institutional data marts, accurate cohort counts have been achieved.

Index cases of *Clostridium difficile* (CDiff) infection were identified, either by the presence of a diagnosis code or by laboratory test results. The first difficulty arises in recognizing resolved CDiff infection: how to differentiate between refractory and recurrent infection. If there is no encounter with CDiff code, laboratory test or relevant medication within eighteen days of date of diagnosis or of positive test result, the infection is assumed to have cleared. Any further infection in 18 to 56 days post index date is recorded as recurrence. Infections later than 56 days are considered new rather than recurrent.

One of the key challenges to CAPriCORN's distributed architecture will be the identification of recurrence across institutions. This challenge has not yet been attempted, but will be among the first studies that the system will address. The cohort is anticipated to be relatively small and the patient cases moving from one institution to another, while at risk of recurrence of CDiff, should be fewer still, so that discovery of such cases will represent success with truly rare events.

Conclusion

Along with ten other CDRNs, CAPriCORN is at the halfway point of its "Phase I" life span and is ready to test its systems with real use cases. The infrastructure was designed to allow for evolution in the data model and increasing complexity of queries in the future. Five submitted cohort studies are currently being processed through stages of the CAPriCORN workflow, and a number of new study proposals are being prepared.

Sustainability of the architecture will be demonstrated through a number of additional research studies that had not been considered at the proposal stage. These studies provide a valuable challenge to CAPriCORN's proposal triage, patient-centeredness, and external researcher engagement workflows.

Acknowledgements

The authors are principal informaticians on the CAPriCORN project. Acknowledgement is due to the overall project PI, Terry Mazany of the Chicago Community Trust for his leadership, and to site PIs: other than those listed as co-authors, Fred Rachman of the Alliance, Raj Shah at Rush, and Brian Schmitt at VA Hines; also Jerry Krishnan who was the original site PI at UIC. Doriane Miller and Madeleine Shalowitz direct the PCAC and have provided great clarity on patient engagement. Jonathan Tobin (NYC CDRN) has been a great supporter with ideas and proposals in the External Researcher Committee. Last but not least, John Collins and Shelly Sital at the Illinois Medical District Commission have provided a consistently high standard of project support.

References

[1] Patient Centered Outcomes Research Institute [Internet]. PCOR. [cited 2014 Dec 22]. Available from http://www.pcori.org/content/research-we-support

[2] Abel N. Kho, John P. Cashy, *et al*. Design and Implementation of a Privacy Preserving Electronic Health Record Linkage Tool in Chicago. (Under review at JAMIA.)

[3] HealthLNK Data Repository [Internet]. [cited 2014 Dec 22]. Available from: http://www.healthlnk.org/

[4] Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, Brown JS. Design of a national distributed health data network. Ann Intern Med 2009; 151:341-4.

[5] PopMedNet [Internet]. [cited 2015 Apr 9]. Available from: http://www.popmednet.org

Address for correspondence

Anthony Solomonides, Center for Biomedical Research Informatics, NorthShore University HealthSystem, 1001 University Place, Evanston, IL 60201, USA. Email: asolomonides@northshore.org

Abel Kho, Division of General Internal Medicine, Northwestern University Feinberg School of Medicine, Rubloff Building 10th Floor, 750 N Lake Shore Drive, Chicago IL 60611, USA. Email: Abel.Kho@nmff.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-589

CA²JU: an Assistive Tool for Children with Cerebral Palsy

Flávio Arthur O. Santos[a], Carlos Augusto E. M. Júnior[a], Hendrik Teixeira Macedo[a], Marco T. Chella[a], Rosana C. do Nascimento Givigi[b], Luciano Barbosa[c]

[a]*Computer Science Department, Federal University of Sergipe, Brazil*
[b]*Speech and Hearing Department, Federal University of Sergipe, Brazil*
[c]*IBM Research - Brazil*

Abstract

This paper presents CA²JU, a hardware/software tool that aims to help individuals with severe speech or language problems in their communication in order to promote their social and digital inclusion. CA²JU is composed of two applications: CA²JU Accelerated, which makes typing faster by suggesting potential words to the user; and CA²JU Illustrated, which automatically converts a sentence of words into a sequence of pictographic symbols, allowing a user familiar with the symbols to verify whether the written sentence is correct. We have implemented, evaluated in a controlled scenario, and deployed CA²JU in a real environment with children with cerebral palsy. In the controlled settings, the results confirm CA²JU Accelerated speed up typing by reducing the number of clicks made by users, and CA²JU Illustrated obtained high accuracy by suggesting the correct pictograms from sentences. In the real scenario, the two use cases show that the children improved their communication and linguistic abilities.

Keywords:

Natural Language Processing; AAC; Web Interface.

Introduction

Communication is a key attribute for a human being. It is fundamental for expressing necessities and desires, enabling access to information, and socialization. Because of that, when a person lacks the power of communication, it is frustrating for her and for people around her. As a consequence, many assistive technologies have been proposed to deal with these limitations such as, for instance, Augmentative and Alternative Communication (AAC) [1],[2],[3]. According to the American Speech and Hearing Association [4]: AAC is a research and educational area of clinical practice for Speech Therapists who aim to "compensate and facilitate, temporarily or permanently for the impairment and disability patterns of individuals with severe expressive and/or language comprehension disorders". AAC allows communicative interactions of individuals with speech impairment, helping them in the process of language building and school interaction, which are fundamental for learning.

Natural language processing can have an important role in AAC [5]. For instance, it can reduce the input necessary to produce a text, increasing the communication speed; it can add semantics to written text to improve its comprehension etc. For instance, [1] described an English language word prediction tool that has a virtual keyboard. Mihalcea and Leong [6] proposed a system for the construction of pictorial representations by sentences. Finally, Nakamura and Zeng [7] described a text-pictogram conversion system to assist the creation of illustrated education for patients. Unfortunately, these and other initiatives focused primarily on the English language, lack proper integration between the software and the needed hardware, and do not provide an updated and intuitive Web interface which actually aids both the patient and therapist to use the software.

This paper presents the assistive tool called CA²JU, a hardware/software solution that aims to improve the communication in Brazilian Portuguese of children with cerebral palsy. The hardware is the interface of the system with the users. It allows the children with disabilities to provide the input to the software. The software is composed of two applications: CA²JU Accelerated and CA²JU Illustrated. CA²JU Accelerated predicts the next words to be typed by the user. It uses a NLP technique known as language modeling to perform this task. We implemented two different approaches for this task: N-gram and HMM. CA²JU Illustrated allows users to select pictograms to create sentences by converting sequences of pictograms into written sentences. To perform this task, it uses different NLP techniques such as Stemming and Name Entity Recognition to clean and add semantics to the sentences.

We evaluated CA²JU in a controlled setting and the results show that: (1) CA²JU Accelerated considerably reduced the number of clicks needed to be performed by the users to write sentences, where HMM showed a superior performance over N-gram; and (2) CA²JU Illustrated demonstrated a high accuracy of suggesting the correct sequence of pictograms from known sentences. We also report two use cases showing the experience of children with cerebral palsy using CA²JU. The child who used CA²JU Accelerated presented a great improvement in her communication and linguistic abilities. Initially, the child produced meaningless texts but, after many sessions using the application, she was able to write sentences with some semantic structure. The other child, who used CA²JU Illustrated, could communicate more effectively and improved his linguistic knowledge after starting to use the application.

The remainder of this paper is organized as follows. In Section 3, we explain CA²JU in detail, showing its input devices, architecture and the techniques used to perform its tasks. In Sections 3 and 4, we present the experimental evaluation of the deployment of CA²JU Accelerated and CA²JU Illustrated, respectively. We conclude Section 5 with our final remarks.

The Tool CA²JU

As we mentioned before, CA²JU is composed of two applications: CA²JU Accelerated and CA²JU Illustrated. In this section, we describe them in more detail.

CA²JU Accelerated

CA²JU Accelerated aims to speed up text input in Brazilian Portuguese of children with cerebral palsy. It does so by using word prediction to suggest to users the next word to input [8]. Figure 1 presents the user interface of CA²JU Accelerated. It contains a virtual keyboard used to input text, and a list of suggested words for the sentence currently being typed.

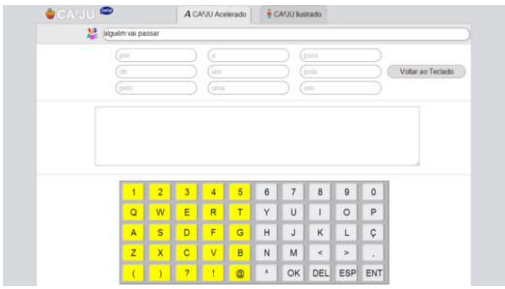


Figure 1 - Interface of CA²JU Accelerated

Language models are probabilistic models commonly used for word prediction. In this paper, we implemented two types of language models: N-gram and Hidden Markov Models (HMM).

Word prediction using N-gram

The joint probability of a given sentence *S* composed by $W_1 \ldots W_n$ words can be calculated using the chain rule:

$$P(W1, W2 \ldots Wn) = P(W1)P(W2 \mid W1) \ldots P(Wn \mid W1 \ldots Wn\text{-}1) \quad (1)$$

Since it is not computationally feasible to keep the whole previous context for all words, we limit the previous context of a word *W* to a fixed number of N-1 words. As a result:

$$P(W_n \mid W_1 \ldots W_{n-1}) \approx P(W_n \mid W_{n-N+1} \ldots W_{n-1}) \quad (2)$$

This is known as the N-gram model [9]. The assumption is that knowing the fixed previous context of *W* is enough to predict the probability of *W* (Markov assumption). For bigrams (N=2), for instance, the context is the immediately preceding word. Using bigrams, the joint probability presented in equation 1 becomes:

$$P(X_1, X_2 \ldots X_n) \approx P(X_1)P(X_2 \mid X_1) \ldots P(X_n \mid X_{n-1}) \quad (3)$$

To calculate the N-gram conditional probability, we use Maximum Likelihood Estimation (MLE). For bigrams, we have:

$$P(W_n \mid W_{n-1}) = C(W_{n-1}, W_n) / C(W_{n-1}) \quad (4)$$

where $C(W_{n-1}, W_n)$ is the number of times the words W_{n-1} and W_n appear together in the corpus, and $C(W_{n-1})$ the frequency of W_{n-1} in the corpus.

Word prediction using HMM

Another way to perform word prediction is using Hidden Markov Models [10]. In this work we have implemented the first order HMM, in which the current state depends only on the previous state and not on all previous states, so we can compare with the word prediction algorithm using N-GRAM, and then decide which algorithm to use. Hidden Markov Models is a temporal probabilistic model, where the state of the process is characterized by a single discrete random variable. In our case, the possible values of the variable representing the state are the parts of speech of the Portuguese language. The model is composed of a set of state variables:

$$X_t = \{X_1, X_2, \ldots, X_N\} \quad (5)$$

A set of evidence variables:

$$E_t = \{E_1, E_2, \ldots, E_K\} \quad (6)$$

A T_{ij} matrix of state transition probability i to state j:

$$T = [P(X_j \mid X_i) * P(X_i)] = \begin{bmatrix} T_{11} & T_{12} & \ldots & T_{1j} \\ T_{21} & T_{22} & \ldots & T_{2j} \\ \vdots & \ldots & \ldots & \ldots \\ T_{i1} & \ldots & \ldots & Ti_j \end{bmatrix} \quad (7)$$

A S_{ik} matrix of state transition probability X_i for E_k evidence variable:

$$S = [P(E_k \mid X_i) * P(X_i)] = \begin{bmatrix} S_{11} & S_{12} & \ldots & S_{1k} \\ S_{21} & S_{22} & \ldots & S_{2k} \\ \vdots & \ldots & \ldots & \ldots \\ S_{i1} & \ldots & \ldots & S_{ik} \end{bmatrix} \quad (8)$$

An array of evidence A_{1K} output variable probability belonging to the state X_i, X_i for each state:

$$A_{1k} = [P(E_1)\ P(E_2)\ \ldots\ P(E_k)] \quad (9)$$

To perform word prediction, we classify the input word according to its grammatical class, Part of Speech Tagging [9], then consult the matrix presented in equation (8) regarding the part of speech resulting, and then get the possible variables of evidence that we have.

CA²JU Illustrated

CA²JU Illustrated allows users to convert a sentence, written in Brazilian Portuguese, into an ordered pictogram representing the meaning of the sentence. Its goal is to help children with cerebral palsy in reading and writing learning, since it allows the child familiar with pictograms to verify whether the written sentence is correct by looking at its corresponding pictogram sequence. In addition, the application provides information about the words such as their part-of-speech tag, synonyms and the entity they belong to (e.g. if it is a place, an organization or a person), which can help them learn these concepts. Figure 2 presents the user interface of CA²JU Illustrated.

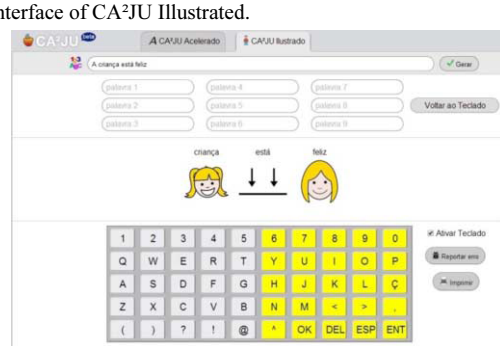


Figure 2 - CA²JU Illustrated's interface

Given a written sentence *S*, CA²JU Illustrated, first applies the RSLP Stemmer algorithm [11] in the words of *S*, so that suffixes of each word are removed. Next, it performs stopword removal, whose goal is to keep only the meaningful words in *S*. Since our original stopword removal, using a list of words, removed meaningful words, we used a Naïve Bayes classifier to identify important parts of the sentences: subject, verb and object (SVO), as well as greeting expressions, which we want to keep in the sentences.

The resulting sentences, with SVO information, are used to map to pictograms. The algorithm searches for the maximum sequence of words in each SVO portion in the sentence that

maps to a pictogram in a given mapping dataset. If no pictogram is found, it does a similar process looking at the root words, identified by the RSLP Stemmer algorithm, instead of the actual words. If still no pictogram is found, a not-found pictogram is presented.

Last, CA²JU Illustrated runs a Name Entity Recognizer (NER) [9] to identify entities such as place, organization and person. From the tagged entities, CA²JU Illustrated is now able to add entity information to their respective pictograms. For the NER detection, we used a logistic regression model, also known as maxent [12]. Maxent performs a linear combination of predictor variables (features) to perform the classification. More specifically, it calculates the probability of that an instance x is member of a class c in the following way:

$$p(c \mid x) = \frac{\exp(W_c x)}{\sum_{c' \in C} \exp(W_c x)} \quad (10)$$

Where W_c is the weight matrix of the features associated to the class c, and C is the set of classes. The instance x is classified in the class with the highest probability.

During training, the algorithm finds the model, represented by W, with the largest entropy from all models that fits the training data. To train the classifier for our task, we used the corpus of the Golden Collection of HAREM [13], which contains 225,000 words.

Buccal Input Device and Virtual Keyboard

The input module is composed of a hardware system for detecting the opening and closing of the mouth, signal conditioning, and the filtering and communication with the computer. The central element of the detector is the module composed of a magnet and a transducer that detects the magnetic field generated by the magnet into electrical signals.

The virtual keyword works in the following way. Initially, its cursor keeps switching between the left half of the virtual keyboard and the right half (see Figure 1). After receiving an input from the buccal device, the cursor vertically keeps changing from line to line on the chosen half. After a second input in the buccal device, the cursor keeps changing horizontally, going over each letter on the chosen line. Finally, the user can select a desired letter by moving the buccal device when the cursor is on the letter.

Figure 3 – Buccal Input Device

Experimental Evaluation of CA²JU Accelerated

It is difficult to assess word predictor softwares because many factors that have influence in the writing time are related to not only the software but also the user (user familiarity and level of disability) and the user environment.

Data Description

We used a generic corpus and personalized corpora for each user in Portuguese for this evaluation. The generic corpus used was Bosque Sintático [14], which has 14,844 words taken from generic news articles. The personalized corpora were obtained from social media conversations, e-mail messages, scientific works, blog posts of 8 users without disabilities. Table 1 presents information about these data.

Table 1 – Information about the personalized corpora

User	Number of words
1	15476
2	2795
3	16793
4	6084
5	33353
6	17426
7	13514
8	113211

Metrics and Test Set

To evaluate the different word prediction strategies, we use the writing speed of the users. For that, we use a metric called Keystroke (KSR) which is defined as follows:

$$\text{Ksr} = (1\text{-KP/KA})*100 \quad (11)$$

KP is the number of inputs on the sensor needed to write the sentence using word prediction, and *KA* is the number of inputs without word prediction. We also used the *HR* metric, which calculates how many times the user selected a word from the suggested list. These metrics were used to calculate how the approaches' perform over one sentence per user (see Table 2).

GRAM and HMM Results

Table 2 presents the results for the N-Gram model and Table 3 the results for the HMM model.

Table 2 - Generic=G, Personalized=P

User	Sentence Size	Ksr in G	HR in G	Ksr in P	HR in P
U1	92	28.26%	3	39.13%	6
U2	66	18.56%	4	48.49%	7
U3	90	29,17%	7	68,06%	15
U4	89	14,61%	3	32,58%	12
U5	104	32,77%	11	33,04%	12
U6	129	14,73%	8	48,84%	14
U7	99	43%	7	51.60%	11
U8	114	15.79%	5	71.28%	18

Table 3 - Generic=G, Personalized =P

Name	Sentence Size	Ksr in G	HR in G	Ksr in P	HR in P
U1	92	34.17%	4	52.28%	10
U2	66	23.29%	4	66.66%	7
U3	90	33.43%	7	70.62%	15
U4	89	19.91%	4	46.34%	15
U5	104	37.58%	12	64.56%	16
U6	129	16,14%	8	54,91%	13
U7	99	48.38%	9	71.75%	14
U8	114	27.81%	8	71.92%	18

The results show that both word prediction models, N-gram and HMM, reduce the number of clicks for all users. In addition, the personalized corpora using HMM outperformed

N-gram in all cases. Another interesting result is that the size of the corpus not necessarily implies the reduction of clicks. For instance, user 5 has a worse result than user 7 but the corpus of user 5 (33353 words) is much bigger than user 7 (13514 words).

Use case of a teenager with cerebral palsy

We now show a successful use case of CA²JU Illustrated with a user with severe quadriplegic mixed cerebral palsy who we in this section call user M. After 74 sessions performed by the Speech and Hearing Department at Federal University of Sergipe, user M presented a great improvement in her writing abilities as we describe next.

In each session, the speech and language therapist worked with user M using CA²JU Illustrated along with linguistic activities. Before using CA²JU Illustrated, user M could not write due to her motor conditions. The first words produced by user M were meaningless, as one can read from the therapist report: "[...] we tested the virtual keyboard and user M wrote disconnected letters and then we asked whether she was writing something or just playing with the keyboard. User M smiled, looking excited and delighted with the possibility of handling the computer by herself "(12 ° Excerpt care; 13.08.12).

After understanding the operation of the system, user M wrote a single word when she wanted to convey some information and as she got more familiar with the system, she started to be able to tell stories using disconnected words, which were understandable for those who were mediating the story creation. Here is what the therapist reported from a story written by user M that she could understand: "[...] The story was that on holiday somebody had traveled to Los Angeles to play football. He had gone with his girlfriend who was a cheerleader. His team whose name was "ESFLAISSDL" had won "ERREST". It was interesting to note that she deleted letters a few times while writing these strange names to convince us that they actually existed "(Excerpt from 32 ° session; 19.11.12).

After a few more sessions, M. felt the need to better structure her writing in order to allow a more effective communication, not only with people close to her, but also to any other person. This is the therapist's repost regarding that: "[…] we realized a progress, a change in the writing of user M, since in previous sessions, user M had never worried about placing the pronoun + verb + noun to build a complete sentence" (Excerpt the 48th session; 11.04.13).

Experimental Evaluation of CA²JU Illustrated

Testing Scenarios

We evaluated CA²JU Illustrated with 10 users without disabilities as follows. First, the users were trained with respect to the pictograms by showing them some pictograms in sentences and their respective meanings. We then converted 10 sentences into sequences of pictograms, hiding the word associated with each pictogram. These sequences of pictograms were presented to the users, asking them to write associated sentences. Finally, we measured the similarity between the correct sentences and the written sentences by the users using the metric presented in Section 4.3.

Dataset and Metrics

The pictogram dataset used in the experiments contains 13,013 colored pictograms from CATEDU [15].

The metric used was the Levenshtein distance [16], which calculates the similarity between two sentences based on the number of character operations (i.e. insertions, deletions or substitutions) needed to change one word into the other. The Levenshtein distance is calculated by the following formula:

$$lev_{a,b}(|a|,|b|) = \begin{cases} \max(|a|,|b|) & , \min(|a|,|b|) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & , otherwise \end{cases} \quad (12)$$

Results

Tables 4 and 5 show the results for the 10 users and the 10 sentences.

Table 4 –Results from sentence 1 to 5

Users	Sent.1	Sent.2	Sent.3	Sent.4	Sent.5
User1	84.31	66.67%	100.0%	75.00%	62.22%
User2	56,00	100,0%	95,65%	83,33%	100,0%
User3	67,92	83,33%	35,00%	71,21%	100,0%
User4	30,00	100,0%	20,00%	20,41%	83,78%
User5	50,00	45,83%	73,91%	46,29%	54,34%
User6	40,00	62,50%	95,65%	65,71%	59,45%
User7	60,00	100,0%	61,76%	86,53%	89,18%
User8	72,00	100,0%	95,65%	42,26%	100,0%
User9	16,00	100,0%	41,66%	56,34%	100,0%
User10	34,00	100,0%	50,00%	25,49%	100,0%
Avg.	42,68	85,83	66,93	57,26	84,90

Table 5 - Results from sentence 6 to 10

Users	Sent.6	Sent.7	Sent.8	Sent.9	Sent.10
User1	81,48	58.82%	76.47%	62.90%	75.00%
User2	95,45	61,76%	82,97%	62,29%	75,00%
User3	98,18	58,82%	75,00%	64,70%	75,00%
User4	89,74	17,65%	72,09%	61,01%	100,0%
User5	41,66	23,52%	64,28%	57,62%	20,00%
User6	80,55	17,64%	54,76%	57,62%	40,00%
User7	94,44	52,94%	100,0%	59,32%	75,00%
User8	98,18	61,76%	95,23%	34,92%	75,00%
User9	62,04	61,76%	88,09%	66,10%	35,48%
User10	75,00	20,58%	100,0%	50,00%	100,0%
Avg.	81,67	43,53	80,89	57,65	67,05

The numbers show that the similarity between the sentences created by the users was fairly similar to the correct ones. For instance, only 2 out 10 users had an average similarity lower than 0.5. These results confirm that CA²JU Illustrated is really able to present the meaning of sentences in sequences of pictograms. Another observation is that for certain sentences the similarity among users varies a lot. This can indicate the users might have different interpretations of the pictograms, and they might need to be disambiguated in some way.

Use case of a teenager with cerebral palsy

User G is a 12-year-old male teenager, diagnosed with severe quadriplegic mixed cerebral palsy. He does not speak, neither studies at a special school. He has weekly speech and language therapy sessions where one of the goals is to teach him how to use CA²JU Illustrated. During the sessions where we used the software, user G understood well its functionalities. He interacted with CA²JU Illustrated by kicking a sensor device (see Figure 4). Using this input device, user G is able to select pictograms to form sentences.

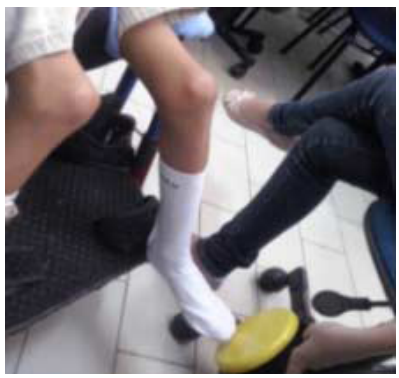


Figure 4 – User G using the device.

The use of CA²JU Illustrated has positively contributed to his communication, improving his linguistic structure. We plan to allow him to use CA²JU Illustrated in other environments such as his home and school.

Conclusion

We presented in this paper CA²JU, our AAC tool composed of CA²JU Accelerated, which helps users with disabilities to type text by suggesting the next word, and CA²JU Illustrated, which helps users to compose sentences by using pictograms.

We evaluated both applications in controlled scenarios, as well as in real ones, with children with cerebral palsy. For CA²JU Accelerated, the results showed word prediction reduced the number of clicks needed to create a text, and its use by a teenager with disabilities helped improve his linguistic skills and communication. The results of CA²JU Illustrated confirm that it creates meaningful sequences of pictograms and also helped a teenager with cerebral palsy to improve his communication.

References

[1] Trnka, K., McCaw, J., Yarrington, D., McCoy, K. F., and Pennington, C. User interaction with word prediction: The effects of prediction quality. p. 34.

[2] BEUKELMAN, D., and MIRENDA, P. Augmentative and alternative communication.

[3] LIGHT, J. Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. augmentative and alternative communication. p. 137–144.

[4] ASSOCIATION, A. S. H. Augmentative and alternative communication. p. 9–12.

[5] Newel, A., Langer, S., and Hickey, M. The role of natural language processing in alternative and augmentative communication. p. 1–16.

[6] Mihalcea, R., and Leong, C. Toward communicating simple sentences using pictorial representations. machine translation. p. 153–173.

[7] Nakamura, C., and Zeng, Q. The pictogram builder: Development and testing of a system to help clinicians illustrate patient education materials. p. 324–330.

[8] ANSON, D. The effect of word prediction on typing speed.

[9] Jurafsky, D., and James, H. Speech and language processing an introduction to natural language processing, computational linguistics, and speech. 2000.

[10] JORDAN, M., PEREIA, Paulo C. BRITTO JR, A. d. S., STELLE, L., and NOHAMA, P. Predição de palavras: Desenvolvimento de uma técnica baseada em Markov.

[11] Orengo, V. M., Buriol, L. S., and Coelho, A. R. A study on the use of stemming for monolingual ad-hoc portuguese information retrieval. Springer. p. 91–98.

[12] BERGER, Adam L.; PIETRA, Vincent J. Della; PIETRA, Stephen A. Della. A maximum entropy approach to natural language processing. Computational linguistics, v. 22, n. 1, p. 39-71, 1996.

[13] Coleção dourada do HAREM, http://www.linguateca.pt/primeiroHAREM/harem.html, 2014.

[14] Corpus bosque sintático, http://www.linguateca.pt/floresta/corpus.html, 2014.

[15] Palao, S. Catálogo de pictogramas coloridos. http://www.catedu.es, Mar. 2014.

[16] Levenshtein, V. I. Binary codes capable of correcting deletions, insertions and reversals. in: Soviet physics doklady. p. 707–710

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-594

Generation of Natural-Language Textual Summaries from Longitudinal Clinical Records

Ayelet Goldstein[a], Yuval Shahar[a]

[a] Medical Informatics Research Center, Department of Information Systems Engineering, Ben Gurion University of the Negev, Beer Sheva, Israel

Abstract

Physicians are required to interpret, abstract and present in free-text large amounts of clinical data in their daily tasks. This is especially true for chronic-disease domains, but holds also in other clinical domains. We have recently developed a prototype system, CliniText, which, given a time-oriented clinical database, and appropriate formal abstraction and summarization knowledge, combines the computational mechanisms of knowledge-based temporal data abstraction, textual summarization, abduction, and natural-language generation techniques, to generate an intelligent textual summary of longitudinal clinical data. We demonstrate our methodology, and the feasibility of providing a free-text summary of longitudinal electronic patient records, by generating summaries in two very different domains – Diabetes Management and Cardiothoracic surgery. In particular, we explain the process of generating a discharge summary of a patient who had undergone a Coronary Artery Bypass Graft operation, and a brief summary of the treatment of a diabetes patient for five years.

Keywords:

Summarization; Natural Language Generation; Temporal Abstraction; Knowledge Representation.

Introduction

Increasingly, large amounts of clinical data accumulate during the process of managing patients, in particular, chronic patients. To make diagnostic and therapeutic decisions regarding an individual patient, physicians increasingly need support to assist them in quickly comprehending the longitudinal clinical course of that patient, preferably a free-text summary of the quality that would be produced by a knowledgeable clinician. A high level of summarization, however, requires more than just enumerating the diagnoses made, the procedures performed, the medications administered, and the results of the laboratory tests taken. An intelligent analysis requires also an explicit representation and application to the data of relevant clinical knowledge.

Visualization is an effective approach to facilitate analysis and presentation of high volume data, specifically when dealing with huge amounts of data [2]. However, other findings showed that graphical representation is not always more effective than other methods, and in the medical domain, clinical decision-making was not necessarily improved by the use of a graphical display. For example, studies have shown that a graphical representation is not always more effective than other methods [3], and that in the case of large and dense graphs, common when high level complex decision making is required, graphical representation has marked disadvantages [4]. Other studies have demonstrated that graph interpretation requires complex cognitive processes [5, 6] and that errors can be made even when interpreting simple graphs [7, 8]. More recent work [9] compared the interpretation of business process descriptions presented in a graphical notation and in an alternative textual notation. Their work has demonstrated an increased understanding of the process when the textual model was read by many different users, while the graphical representation showed an increased understanding only in trained readers. A study in the medical domain [10] has shown that when dealing with large volumes of complex clinical data, a textual presentation might even be advantageous over a graphical one, for the purpose of certain clinical tasks.

The *Natural Language Generation* (NLG) task deals with the generation of natural language from a machine-language form input [13]. Although the NLG task has been implemented in different domains [14], in most of the implemented systems the data are relatively well-defined, not requiring advanced data analysis techniques. Furthermore, in the case of the summary of small data sets, only brief summaries are produced, which significantly reduces the complexity of many NLG tasks. In the medical domain, existing NLG systems are far from optimal [14]. A more recent NLG system [15], which focuses on decision support, also performs temporal data abstraction; however, the abstraction process is relatively simple, and does not consider different contexts when determining the importance of an event. In addition, it considers only short and pre-defined periods of data. Dealing with different periods of time, especially longer periods, as is common in chronic-disease patients, requires additional temporal-information handling techniques.

Our objective is to provide an intelligent verbal (free-text) summary of electronic patient records. These records include time-stamped data that have accumulated during an extended time period, such as during hospitalization, or over years of medical care. Such summaries might help care providers in their daily tasks, and support several decision-making processes.

In a preliminary study [1], we had proposed an architecture that supports a process of transforming longitudinal data into an intelligent, concise, text-based summary. In the current study, we have implemented the proposed architecture as a prototype system, the **CliniText** system. In this paper, we describe the inner workings of the CliniText system, using detailed examples of how each module contributes to the process of automatically transforming time-based data into a free-text summary. To be concrete, we will focus on the conversion of a typical case in the electronic cardiothoracic surgery MIMIC-II public database, into a free-text discharge summary, and on the generation of a summary of the management of a diabetes patient, given his five year electronic record in a collaborating health maintenance organization (HMO).

The input to the CliniText system includes longitudinal patient data and domain-specific knowledge. The output of the system

is a condensed textual summary of the patient's data. We opted to use only numeric or structured input data, because that is the input type usable by the temporal-abstraction method that we use to interpret the data. Including textual segments would require methods such as natural language processing, to first analyze and understand the text. Furthermore, the accuracy of such methods is far from 100%, but the requirement for highly accurate output is of major importance in the medical domain.

Methods

Our input data arrives from a time-oriented clinical database. The required medical knowledge is specified through a graphical knowledge specification tool called Gesher [16] using a Temporal Abstraction Knowledge (TAK) schema we developed. We define both *declarative* (e.g. What is mild-anemia in a pregnant women) and *procedural* (e.g. How to manage the patient, by administering two drugs in parallel) knowledge, necessary in the abstraction process. The knowledge is also exploited in other modules of the system.

We implemented the CliniText system using a framework composed of several modules, each performing a specific task. The architecture of the system, composed of six main modules, is displayed in Figure 1.

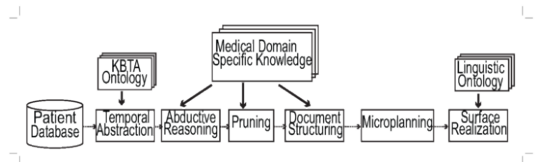


Figure 1. The architecture of the Clinitext System. Knowledge is used by the temporal abstraction module to abstract the data. Additional knowledge is used in further modules to perform abduction, pruning and document structuring of the data.. Dashed horizontal arrows = control/data flow; Bold vertical arrows = knowledge flow

The flow of the data between the components defines the workflow of the process; longitudinal raw data coming from the time-oriented database is abstracted by the *Temporal Abstraction* module, adding abstract data to the original raw data. Additional data is inferred through abduction in the *Abductive Reasoning* module. Raw, abstract and abducted data are then pruned in the *Pruning* module, responsible for selecting the important and domain-relevant data, leaving only the information that will appear in the final text. The *Document Structuring* module structures these data, determining the order they should appear in the final text. The *Microplanning* component groups and prepares the structured data to the expected format by the *Surface Realization* module, which finally realizes the data into the final text.

We shall now describe each component and its role in the process, using two example patient records: one who had undergone a coronary artery bypass graft operation, and one who had been followed for five years due to diabetes.

Temporal Abstraction (TA)

The TA step is responsible for the performance of the abstraction of time-oriented data. In the implementation of this module, we used a variation of the IDAN temporal-abstraction mediator [17], which implements Shahar's *Knowledge-Based Temporal-Abstraction* (KBTA) method [11]. Our input to the TA module, is time-stamped raw concepts (e.g., red blood-cell [RBC] values).

Abductive Reasoning

Not all important concepts, such as events or actions performed, can always be found in the input data. However, in some cases, these events can be inferred by a process of *abduction* from the existing data, with a high probability, albeit not necessarily with complete certainty. For example, knowing that a Swan-Ganz catheter (SG) is used to measure the pulmonary artery systolic/diastolic pressure (PAP) concept, even though the event of inserting a SG (SG-in) does not appear in the database we can hypothesize that the event SG-in occurred through an instance of the PAP concept in the database.

We can say that knowing that concept C_1 *enables* the existence of concept C_2, within a certain time, Δt, allows, given the existence of C_2 in the database, with a timestamp t_2, the abduction of C_1 with a timestamp between t_2-Δt and t_2, as shown in Figure 3. In general, of course, the abduction process involves uncertainty. Formally, assuming for simplicity only one enabled or associated item per concept, we compute ahead of time, for every *i*, $P(C_i[t_2-\Delta t_i] \mid C_2[t_2])$: The probability of each concept C_i, given the concept C_2, t_2 (timestamp of C_2), $Prior_i$ (a-priori probability of each related concept C_i), $P(C_2[t_0+\Delta t_i] \mid C_i[t_0])$ (the probability of C_2, given C_i) and Δt_i (the [past] interval of time during which C_i can hold, given C_2. For simplification, we consider only a uniform probability measure). The computation is then a simple temporal-extension of Bayesian diagnostic problem solving [18].

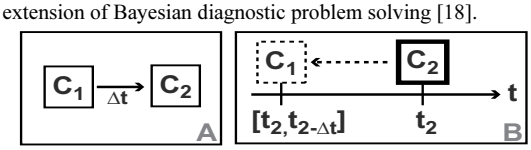


Figure 2. An abduction example. A: The knowledge: concept C1 is followed by concept C2 within a period of Δt. B: The abduction process: Concept C2 at t2 allows the abduction of concept C1 with a timestamp between t2 and t2-Δt

From a knowledge representation point of view, the abduction knowledge typically links declarative data items (e.g., measurements) in the patient's record, to procedural data that enabled the data's existence, although other combinations are possible. An example of linking two instances of procedural data is the abduction from the event "syringe-insertion," of the event "skin-sterilization" 5 to 10 seconds before; an example of abducting a declarative datum from a procedural one is that administration of antibiotics in certain contexts suggests the existence or the suspicion of infection. The relationship between the concepts can be either causal (A causes B) or associative (A typically occurs before B).

To perform abduction, we need to explicitly represent abduction knowledge in the KB. For example, we need to specify beforehand in the procedural knowledge for each event (e.g. SG-in) what are the declarative concepts (e.g. PAP) that it enables or is associated with, and what is the probability/strength of this relation (e.g., are there other events that could also generate this declarative concept?). For example, the event SG-in (whose prior probability is quite low) enables PAP with P(PAP|SG-in) = 0.95. Furthermore, in the ICU, it is the *only* method used to measure PAP. Thus, regardless of the low prior probability and the association strength, in the ICU context, P(SG-in|PAP) = 1.

In our case, we enable the inference of the existence of a procedural action (e.g. SG-in) by the existence of a declarative concept (e.g. PAP). Since we deal with the medical domain, we abduct only concepts that have a probability that is close to certainty, to avoid as much as possible the generation of

incorrect summaries. We denote events abducted from declarative concepts as *inferred interventions*.

Table 1 shows several interventions, inferred through abduction, for a cardiothoracic surgery record.

Table 1. Declarative concepts used to generate "inferred interventions". An intervention can have modifiers (e.g. "inserted" or "removed"), according to the value of the raw concept from which it was deducted (e.g., certain PAP values indicate non-measurement)

Raw concept(s)	Inferred inter-ventions	# instances
[PAP S/D]	SG catheter	1
[CVP], [A.BP]	Arterial Line	3
[Airway Size], [ETTMark]	Artificially breathing	1

Pruning

After enriching our original data with abstractions and inferred interventions, we need to select what information is important and what should be removed to avoid overloading the user.

We prune the data by using general heuristics and several "independent" parameters (e.g., text length, detail level, profiles, time-period, etc). We use *pruning-heuristics*, which define which data should be pruned, and *maintaining heuristics*, defining which data instances should be maintained (and which override the pruning-heuristics).

The current maintaining heuristics include: (1) data instances that must be described; these are determined according to the domain, although certain events appear in every domain, such as death of the patient; and (2) maintaining extreme raw data values. In this case, even though they have an abstract value, we might want to maintain the raw data also.

Document Structuring

After considering which data will appear in the final text, we define how these data will be presented. In the *Document Structuring* (DS) module it is decided how much information will be expressed in each phrase, in which order the facts will appear, how they will be organized into paragraphs, etc. The structure of the final text varies according to the output format expected as well as the domain.

We use two approaches for the DS: a *top-bottom* approach, which defines the main structure of the final text (e.g. division of paragraphs), and a *bottom-up* approach, which groups and combines events until all the events are linked together, by using appropriate discourse relations.

The output of the DS module resembles a tree. We prepare ahead of time, for each domain, a "stub" of the tree which includes the main textual segments in that domain, in a top-bottom approach. We build up from the pruned data a set of intermediate nodes that we try to link to the most appropriate node in the pre-prepared stub, in a bottom-up fashion. We call the leaves of the tree Document-Messages and the inner-nodes of the tree Document-Blocks. The tree has some parts that are more structured, part of the "stub" pre-defined, and others that are more flexible, with their structure affected and defined by the patient's data. An example of the trees used in both, the cardiac and diabetes, domains can be seen in Figure 3.

Document-blocks can be rigid or flexible. For example, in the cardiac domain, the Document-blocks: "*Admission Details*", "*Operation Admission*", "I*CUOutDischarge*", and "*Discharge Details*" have a rigid structure, which was mostly defined previously, while the blocks "*Operation Procedure Routine*" and "*Operation Recovery*" are more flexible and their context will be determined and modeled based on the patient's actual instances of data associated with them. *Document-Messages* can be generic, such as in the case of Drug-Message, Intervention-Message, or Data-Message, or specific, such as Admission-Message, PrevDiseases-Message, and ICUAdmission-Message.

Structured-blocks will be populated with the specific Document-Messages associated with it. In the cardiac domain for example, the structured section "Admission Details" will be populated with Document-Messages: Admission-Message and PrevDiseases-Message. Each specific Document-Message has specific data associated with it. Less-structured blocks may contain further Document-Blocks as well as generic Document-Messages. The creation of further Document-Blocks is affected by contexts that the data may generate. For example, the less-structured Document-block OperationRoutine will have further Document-Block children generated by the contexts associated with a specific operation event. Note that in general, the TA process applied by the KBTA method in the first step already generates (in addition to the temporal abstractions) all domain-specific contexts as well.

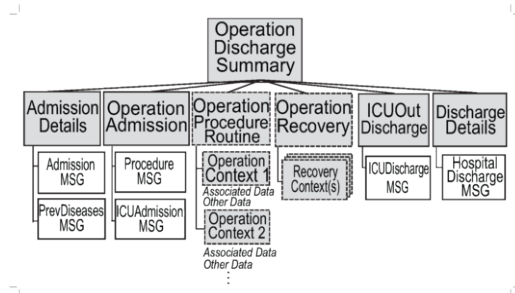


Figure 3. The document structuring tree in the case of the cardiothoracic surgery domain. The Document Structuring module structures the final text into a tree. The tree defines how the data will be presented – the order, the paragraphs division and also the relation between events. (Gray rectangles are Document-Block objects, white rectangles are Document-Message objects. Document-Blocks can contain one or more Document-Messages or Document-Blocks.) Dashed-lined blocks represent flexible segments whose structure is defined by the data. Non-dashed blocks represent rigid segments of the text, whose structure was pre-defined

The structure of the Document Structure tree establishes the structure of an operation discharge summary for any given patient that has gone through an operation. To produce a Discharge Summary specific for a given patient, we need to populate the tree with the patient's data. After populating the tree, by traversing the tree in a Depth First Search, we generate the order of the events in the final text.

Microplanning

In the *Microplanning* step, we (1) define the words and syntax structures that are used by the final text generator, (2) perform aggregation of the data, and (3) generate referring expressions.

In the aggregation process, we define how much information will be expressed in each sentence. In the aggregation process, the semantic order has preference over the chronological order that was kept until now. Taking into consideration semantic considerations, we group together messages according to their type (drug, intervention or data message types).

In the case of drug messages, for example, we group together different (drugs) with the same value. For example, instead of:

"drug X was started, drug Y was started and drug Z was stopped" we will have: “drug X and Y were started and drug Z was stopped". Note that messages are grouped only if they occur within the same context, although not necessarily during the same time.

The exact format of the Microplanning output depends on the expected format of the step following it, Surface Realization.

Surface Realization

Surface Realization is the module responsible for generating the final text. Consequently, this step is language dependent, and observes the grammar rules of the language chosen. We opted to implement the Surface realization module using the existent Java module “SimpleNLG” [19].

The realization of the text is done per section. Sections were defined within the tree structure, during the DS step. When the tree is flattened, the sections are kept and the flattened list of messages is divided into these sections. Each section is realized separately and in the same order it appears in the flattened list. One section might have sub-sections (generated by contexts). For example, the OperationRoutine section has two contexts: “Early-postCABG day” and “Late-postCABG day”. Each context (or sub-section) is also realized separately.

Each section has a relationType (defined in the DS step) defining how messages within the section are related. For example the "sequence" relation defines that they should be realized one after the other, while the “causal” relation defines a cause-effect relationship between the messages.

Preliminary Evaluation

In the cardiothoracic surgery domain, we used data from the Multiparameter Intelligent Monitoring in Intensive Care II database (MIMIC-II) [20]. There are two types of data in the MIMIC-II database: clinical data (numeric and textual) and bedside monitor waveforms. We used only the first type, and mainly numeric data. Textual data were used only to compare our generated text to the original discharge summary. We focused on data of patients in the cardiac domain, more specifically on patients that went through a CABG procedure.

Our current evaluation was purely technical, judging the feasibility of the overall process, and comparing the generated text to the MIMIC texts, without involving any users yet.

The diabetes dataset, provided as part of collaboration with an Israeli health-maintenance organization, contains data that were collected each month from 2002 to 2007. The dataset contains six measurements, including hemoglobin A1c, blood glucose level, and cholesterol values, and the medications the patients purchased, including oral hypoglycemic medications, cholesterol reducing statins, and beta blockers, with a mean of 39.5 records per patient. In this dataset, there was no corresponding human text to compare our summary to; nevertheless, the generated text allows one to judge the general quality of the text (readability, soundness, etc.).

Results

Table 2 compares the generated text summary for the cardiothoracic example record with the original human-generated discharge summary found in the public database we used. Part of the original summary could not be reproduced, since there was no data in the database that would support it. Presumably, such parts were written using some external data source, or perhaps textual progress notes.

Table 2. A comparison between the human- and CliniText-generated text, in the cardiothoracic surgery domain. Striked-through text between brackets [text] indicates text parts that cannot be generated, i.e., without supporting electronic data. Bold text on the right (left) column indicates data that were not explicitly mentioned in the other column, although they could have been

The human discharge summary	**The CliniText discharge summary**
(1) “On **3285-5-17**, the patient was brought to the Operating Room for a redo coronary artery bypass grafting [~~times two and aortic valve replacement.~~] (2) The patient [~~tolerated the procedure well and~~] was transported intubated to the PACU **in stable condition** on a Levophed drip, Milrinone drip, and a Propofol drip. The patient had atrial wires and a chest tube. (3) On postoperative day #1, the patient was continued on his Levophed drip and Milrinone drip over night. The patient was extubated over night and was on a nasal cannula of 4 Lungs: **oxygen saturation 95% on room air**. The patient's **vital signs were otherwise stable**. (4) The patient had a post-operative hematocrit of 26.8; otherwise laboratory values were all within normal limits. [~~The patient was encouraged to be out of bed.~~] Drips were weaned. [~~The patient was started on his Plavix.~~] (5) On postoperative day #2, the patient was off all drips, [~~was started on Aspirin 325 per day and Captopril 6.25 t.i.d.~~] […] (6) The patient was transferred to the floor .	*(1) “On the 15/05/2005 Mr.97, a married, 82 years old patient, was admitted to the hospital. The patient's previous history includes peripheral_vascular disease(s). (2) On the 17/05/2005 the patient went through a "CABG" operation and was admitted to the ICU. (3) In the first hours following the CABG procedure, propofol, levophed and milrinone were started.* ***Swan-ganz catheter, arterial line, chest tube, atrial wires and vent wires were inserted,*** *artificially breathing was started,* ***heart rhythm was Paced.*** *The patient's hematocrit,* ***arterial bp and spo2 were Low,*** *temperature was Afebrile and his urine source was through an external catheter. (4) Later on, propofol was stopped. O2 delivery device was Nasal Cannula, heart rhythm was Normal sinus* ***and blood transfusion was given.*** *The patient's O2 Flow State* ***was 3*** *then 4 ; hematocrit* ***and urine output*** *were Low. (5) On post-oper day #1, the patient's* ***Arterial BP State remained Low*** *and Hematocrit State remained Low ;* ***o2 flow was 2****. Levophed and milrinone were stopped. […] (6) The patient was discharged from ICU on 21/05/2005. The patient was discharged from hospital on 23/05/2005.”*

Below we bring the text generated from the longitudinal example record in the diabetes domain (this particular database did not include human-generated text for comparison):

“Mr.1111 is a male diabetes patient in the age group: 75-84. He was monitored and treated in our clinic between Aug-2002 to Jan-2004. His Creatinine_State and Albumin-State level were normal from Jun-2002 to Jan-2004. To decrease the bloodPressure the patient received calcium channel blockers, beta blocking agents, ace inhibitors and thiazides. Due to infection, the patient received sulfonamides. To decrease the glucose the patient received metformin. Glucose decreased from Jun-2002 to Jan-2004 from grade_3 to grade_2...”.

Judging by examples we have gone through, the overall process is quite feasible and shows definite promise, but we have learned from our current preliminary experiments that much additional domain knowledge needs to be added.

Discussion

In the cardiothoracic surgery domain, it can be noted that in the CliniText summary text, there are data items that do not appear in the original text. If some of these data are redundant,

we could prune them by defining additional heuristics in the pruning step, for example.

Another interesting point to be observed is that in the original text the levels of abstraction varies as compared to the levels of abstraction that appear in the CliniText summary, in which we always strive to have a uniformly high level of abstraction (e.g. "hematocrit = 22.3" vs. "hematocrit state = Low"). Allowing the user to interactively navigate into the raw concepts that the abstraction was abstracted from can provide them with lower levels of abstraction when required. We are currently adding this capability.

Note that the original text, produced by humans, may contain wrong information. In our example (Table 2), it is stated in paragraph #6 that the patient was discharged from the ICU, and in paragraph #7 the error is corrected by saying that the patient was actually kept in the ICU due to bed availability. Such errors cannot occur in the CliniText summary, if the data are correct. On the other hand, when there are erroneous data, we cannot avoid generating an incorrect text. An additional layer of data validity checks could be added to the process to clean the data before the text starts to be generated.

One of the advantages of having a discharge summary text being produced automatically is that the output format is structured and does not depend on subjective factors (e.g. how tired is the physician writing it, or how many discharge summaries she already wrote). Furthermore, the information to be included or omitted has been defined by objective consistent criteria, resulting in an objective data-and-knowledge-based summary. Finally, adding several surface realization modules might enable the future generation of multi-lingual summaries of the same record.

The knowledge acquisition step when applying the system to a new medical domain is an essential step of the process. Although the representation of the knowledge in our knowledge-representation language allows reuse and easy update of the knowledge, and is supported by a previously-evaluated graphical medical knowledge-engineering tool [16], it still requires a collaboration between a medical expert, who provides the medical knowledge, and a knowledge engineer who converts the knowledge into knowledge instances.

The main purpose of building the described prototype was to demonstrate the feasibility of creating a system that can generate a complex knowledge-based textual summary of arbitrarily long time-oriented clinical data, in different domains. We showed that it is possible to produce a readable textual summary, and include the main events and data, using domain knowledge. We intend to evaluate the CliniText system regarding the quality of the generated text and functionality of the summary for relevant pre-defined tasks.

We described a new generalized architecture for generation of free text summaries of longitudinal clinical data. In our system, the input data are allowed to be (1) heterogeneous, which makes the NLG task significantly more complex, (2) of high density, which means that the summaries will not be brief, and (3) longitudinal, with unlimited duration. Unlike the approach we took, most existing approaches are not based on the use of complex knowledge specific to the application domain, and thus cannot automatically create meaningful domain-specific interpretations. Furthermore, these systems cannot decide what data or interpretations are potentially redundant by using a robust domain-specific knowledge base and a formal interpretation theory. Moreover, existing systems do not have the capability for interactive exploration of the resulting text summaries at various domain-specific, semantically meaningful levels of abstraction, a capability for which our system provides the infrastructure.

Conclusion

An architecture such as implemented in the CliniText system can be used to summarize longitudinal clinical data using rich domain knowledge. Further evaluation is needed to assess the quality and functionality of the generated text.

Acknowledgements

Support: EU 7th Framework project No. 287811, MobiGuide.

References

[1] Goldstein A, and Shahar Y. A Framework for Automated Knowledge-Based Textual Summarization of Longitudinal Medical Records. KR4HC workshop. Tallinn, Estonia, 2012.

[2] Combi C, Shahar Y, and Keravnou-Papailiou E. Temporal Information Systems in Medicine. Springer: New York, 2010.

[3] Langan-Fox, et al. Effects of advance organizers, mental models and abilities on task and recall performance using a mobile phone network. App Cognitive Psyc 2006: 20(9), 1143-65.

[4] Huang W, Eades P, and Hong SH. Beyond time and error: a cognitive approach to the evaluation of graph drawings. Proc. Conf. Beyond time and errors: novel evaLuation methods for Information Visualization, 2008: pp. 3.

[5] Mayer RE. Comprehension of graphics in texts: An overview. Learning and Instruction 1993: 3(3): 239–45.

[6] Shah P, Hegarty M, and Mayer RE. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. J Ed Psyc 1999: 91(4): 690.

[7] Mevarech ZR, and Kramarsky B. From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. Ed Stud Math 1997: 32(3): 229–63.

[8] Shah P, and Hoeffner J. Review of graph comprehension research: Implications for instruction. Ed Psych Rev 2002: 14(1): 47–69.

[9] Ottensoosera A, et.al. Making sense of business process descriptions: An experimental comparison of graphical and textual notations. J Systems and Software 2012: 85(3): 596–606.

[10] Portet F, et. al. Automatic generation of textual summaries from neonatal intensive care data. Artif Intell Med 2007: 227–36.

[11] Shahar Y. A framework for knowledge-based temporal abstraction. Artif Intell 1997: 90: 79-133.

[12] Lavrač I, et al. Intelligent data analysis for medical diagnosis: using machine learning and temporal abstraction. AI Com 1999: 11: 191–218.

[13] McDonald DD, and Bolc L. Natural Language Generation Systems. New York: Springer-Verlag, 1988.

[14] Huske-Kraus D. Text generation in clinical medicine-a review. Meth Info Med 2003: 42: 51–60.

[15] Portet F, et al. Automatic generation of textual summaries from neonatal intensive care data. Artif Intell 2009: 173: 789-ci6.

[16] Hatsek A, et al. A scalable architecture for incremental specification and maintenance of procedural decision-support knowledge. Open Med Info J 2010: 255-77.

[17] Boaz D, and Shahar YA. distributed temporal-abstraction mediation architecture for medical databases. Artif Intell Med 2005: 34(1): 3-24.

[18] Owens D, and Sox H. Biomedical Decision Making: Probabilistic Clinical Reasoning, In: Biomedical Informatics (Shortliffe and Cimino, eds.), 2006; pp. 80-129.

[19] https://code.google.com/p/simplenlg/

[20] http://www.physionet.org

Address for correspondence

Ayelet Goldstein - gayelet@post.bgu.ac.il

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-599

Classification of Contextual Use of Left Ventricular Ejection Fraction Assessments

Youngjun Kim[a,c], Jennifer Garvin[b,c], Mary K. Goldstein[d], Stéphane M. Meystre[b,c]

[a] *School of Computing, University of Utah, Salt Lake City, USA*
[b] *Department of Biomedical Informatics, University of Utah, Salt Lake City, USA*
[c] *VA Health Care System, Salt Lake City, Utah, USA*
[d] *VA Palo Alto Health Care System, Palo Alto, CA, and Stanford University, Stanford, CA, USA*

Abstract

Knowledge of the left ventricular ejection fraction is critical for the optimal care of patients with heart failure. When a document contains multiple ejection fraction assessments, accurate classification of their contextual use is necessary to filter out historical findings or recommendations and prioritize the assessments for selection of document level ejection fraction information. We present a natural language processing system that classifies the contextual use of both quantitative and qualitative left ventricular ejection fraction assessments in clinical narrative documents. We created support vector machine classifiers with a variety of features extracted from the target assessment, associated concepts, and document section information. The experimental results showed that our classifiers achieved good performance, reaching 95.6% F_1-measure for quantitative assessments and 94.2% F_1-measure for qualitative assessments in a five-fold cross- validation evaluation.

Keywords:

Heart Failure, Ventricular Ejection Fraction, Medical Informatics, Natural Language Processing

Introduction

Ejection fraction is a measure of the percentage of the blood in a heart ventricle that is expelled during contraction. Knowledge of the left ventricular ejection fraction (LVEF) is critical for the optimal care of patients with heart failure. Multiple life-prolonging treatments for patients with heart failure (HF) have been demonstrated in randomized trials that enrolled patients with an LVEF below 40%. The LVEF is needed to determine if patients will benefit from treatment and ensure that these treatments are prescribed to the appropriate patients.

The U.S. Veterans Healthcare Administration (VHA) Consortium for Healthcare Informatics Research (CHIR) Translational Use-Case Project for Ejection Fraction (TUCP-EF), and the Automated Data Acquisition for Heart Failure (ADAHF) research project aimed at the automated extraction of LVEF mentions, LVSF (left ventricular systolic function) mentions, and their associated quantitative and qualitative assessments, as described in the following examples:

- LVEF mentions (e.g., “left ventricular ejection fraction”, “VISUAL ESTIMATE OF LVEF”, “EF”)
- LVEF quantitative values (e.g., “~0.60-0.65”, “0.45”, “50%”)
- LVEF or LVSF qualitative assessments (e.g., “NORMAL”, “mildly decreased”, “SEVERE”)
- LVSF mentions (e.g., “Global left ventricular systolic function”, “systolic dysfunction”, “LVSF”)

When determining the current and most reliable LVEF assessment in a clinical note, identifying the contextual use of these assessments is important. For example, those that are mentioned as recommendations or historical values can be ignored or used for historical reference for longitudinal evaluation, and knowing whether an assessment was a summary, interpretation or a measurement can be used to prioritize the values used for document-level LVEF assessment. For this study, we focused on classification of the contextual use of LVEF quantitative and qualitative assessments into five different categories: summary, interpretation, technical measurement, recommendation, or past finding.

To our knowledge, this is the first attempt at LVEF assessment contextual use classification in HF, however several studies have focused on classification of similar specialized purpose, such as tumor status [1], lung cancer stages [2], and medication prescription status classification [3]. More for local context recognition and analysis, several Natural Language Processing (NLP) systems have been developed that focused on the negation or other assertions of medical concepts. For negation classification, rule-based systems like Negfinder [4] and NegEx [5] have been introduced. They used regular expressions with trigger terms to determine whether a medical term was negated. In BioNLP-2009 [6] and CoNLL-2010 [7-8] Shared Tasks, detecting negations (and their scope) in the natural language text was the focus. Kilicoglu and Bergler [9] compiled negation cues from the corpus and detected the negation using dependency-based heuristics. Morante et al. [10] implemented two stages of negation scope detection: sentence- level classification and phrase level with memory-based learning.

Assertion classification was the focus of the 2010 i2b2 NLP challenge [11]. The defined task consisted in choosing the status of medical problems by assigning one of the six categories: present, absent, hypothetical, possible, conditional, or not associated with the patient. Several studies then used this challenge data and showed that machine learning approaches [12-14] performed better than handcrafted rule-based systems.

In the following sections, we will describe the methods we used for context classification and present our experimental results with feature contribution analysis.

Materials and Methods

We approached the contextual use classification task as a supervised learning problem. The classifier is given a LVEF assessment as input and must assign one of the five context categories as output. We created Support Vector Machine (SVM) classifiers with a variety of lexical features extracted from the target assessment, associated concepts, and section titles.

The TUCP-EF corpus of clinical notes used for this task consists of echocardiography reports, radiology reports, and other note types in the VHA developed using the Text Integration Utility (TIU) software [15]. These clinical notes were manually annotated for LVEF and LVSF mentions and assessments along with the contextual use of these assessments. We defined each category of contextual use, as explained below:

Contextual Use of Assessments

Summary

Any assessment is appearing in the summary of important findings in the study. Summary is the short and concise briefing of the study, as the contrast to the usually detailed and lengthy interpretation. Examples:

- FINAL IMPRESSION: "Normal" LV function
- EF "55-65%".

Interpretation

Any assessment generated based on a clinician's synthesis of the echocardiography machine metrics and reading of the ultrasound images using expert clinical judgment. This assessment usually appears in the body of the echo report as detailed findings or impressions. Examples:

- Systolic function is "normal" with estimated ejection fraction "60%".
- LVEF appears "mildly reduced" ("35-40%").

Technical Measurement

Any metrics that appear to be taken directly from the echocardiography machine readings. This indicates that the assessment is calculated using various algorithms from the technician's measurement. Examples:

- Ejection fraction is calculated at "42%".
- LVEF (Teichholz): "33%"

Recommendation

Any assessment generated as part of decision support messages or reminders. This is not an assessment associated with the actual patient, but rather serves as a recommendation or instructional guidelines. Examples:

- Please contact primary care provider should patient's ejection fraction fall under "40%".
- If "severe" systolic dysfunction is observed in studies with limited visual, please repeat the study within - two days.

Past Finding

Any assessment from a previous echocardiography study. This reflects the patient's past LVEF assessments and is not the value estimated from the current study. Examples:

- No significant change from a previous study ("50-55%").
- Compared to the last echo,the systolic function has reduced from "normal" to mildly decreased.

Data Description

The TUCP-EF corpus includes 3,060 clinical notes from three different components of the VistA electronic health record files [15]: echocardiography laboratory reports (1140 reports, from 16 VHA medical centers sampled at random), the radiology package (720 reports, from 5 medical centers sampled at random), and TIU (1200 reports, from 17 sites sampled at random). Among these 3,060 reports, 1,465 reports contained at least one of our concepts of interest (i.e., LVEF or LVSF mentions or assessments) and were selected for this project. 2,185 quantitative assessments and 1,278 qualitative assessments were manually annotated. The distributions of contextual use categories are displayed in Table 1. No qualitative assessment was annotated as technical measurement, recommendation or past finding.

Table 1 – Distribution of contextual use categories

	Quantitative		Qualitative	
Contextual use	**Count**	**%**	**Count**	**%**
Summary	944	43.2	811	63.5
Interpretation	927	42.4	467	36.5
Tech. measurement	296	13.6	0	0.0
Recommendations	10	0.5	0	0.0
Past finding	8	0.4	0	0.0
All	2,185	100.0	1,278	100.0

Two of the contextual use categories (Summary and Interpretation) accounted for nearly 85% of the quantitative assessment instances in the data set, while the other three classes were relatively infrequent.

For qualitative assessments, a heart failure expert identified terms used to describe the left ventricular function and provided quantitative EF ranges to correspond to each term, as shown in Table 2. To enhance the consistency of assessments, we normalized and grouped each qualitative assessment into six ranges (<30%, 25-35%, 30-40%, 35%-45%, 40-50%, and >50%) and used a list of assessment terms for this purpose. This numerical range was assigned when the assessment contained one of the assessment terms.

Table 2 – Qualitative assessment ranges and terms

Range	Terms
<30%	Severe
25-35%	moderate to severe
30-40%	Moderate
35%-45%	mild to moderate
40-50%	Mild
>50%	normal, borderline low, lower limits of normal, hyperdynamic, preserved

We also normalized each quantitative value with the same ranges. To assess the consistency of assessments, we compared each pair of assessments in a document with the normalized ranges. 73 clinical notes (about 5% of the corpus) contained assessment pairs that did not agree with each other (i.e., their ranges were different), even when not considering the pairs with overlapping ranges (e.g., one assessment with 25-35% and the other with 30-40%.)

Contextual use classification is crucial to determine which assessment to use to determine the document-level LVEF, especially when multiple assessments have different ranges. When selecting the document-level LVEF, recommendations, and past findings were excluded, and the following order was used to prioritize the assessments:

summary (quantitative) > summary (qualitative) > interpretation (quantitative) > technical measurement (quantitative) > interpretation (qualitative)

In the next section, we describe how the feature vectors were extracted from clinical notes and which feature set was utilized for contextual use classification.

Methods

We built a NLP information extraction application with two pre-processing components and SVM classifiers, as depicted in Figure 1. Pre-processing includes a tokenizer and a section detector.

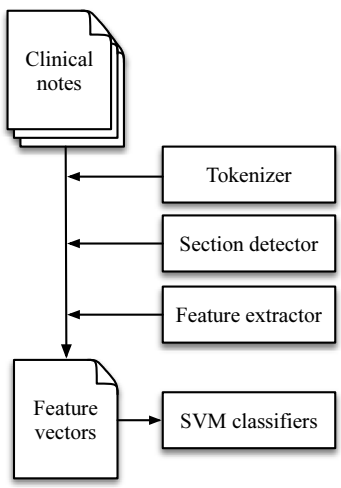


Figure 1 – Application architecture for feature extraction and classification of contextual use

The tokenizer is based on regular expressions and splits the text in groups of alphanumerical characters separated by white space characters. For clinical note section detection, we randomly selected 150 documents (about 10% of the corpus) and compiled the list of section headers related with the concepts of interest. A phrase was detected as a section header when it contained the following keywords: *impression*, *summary*, *findings*, *conclusion*, *assessment*, *interpretation*, *results*, and *note*.

Machine Learning Classification Features

Features extracted from pre-processors are listed below:

- Lexical: A bag-of-words representation of lexical features included the assessment term itself, uni-grams and bi-gams of five words preceding it, and five words following it.
- Concept position: We created two binary features to represent whether the occurrence of the assessment was the only one in the document, and whether any preceding assessment existed before the target assessment in the document. We also defined a feature representing the assessment location normalized by the document length.
- Related concept: We created features for mentions related to the target assessment. We captured the nearest LVEF or LVSF mentions. The mention term itself, uni-gram and bi-gams of five words preceding it, and five words following it were used for this set of features. We also computed the distance by counting the words between the target assessment and the nearest mention.
- Section: We detected the phrases containing the section header keywords listed above. Titles of the nearest section, the nearest previous section, and the nearest following section were used as section features.

Contextual Use Classification

We trained two multi-class SVM classifiers, one for quantitative assessments, and the other for qualitative assessments, both using the LIBLINEAR software [16]. The quantitative assessment classifier assigned one of the five contextual use categories: summary, interpretation, technical measurement, recommendations, and past finding. Due to the absence of annotations in all five categories, the qualitative assessment classifier only made a binary decision between summary and interpretation categories.

Experimental Results

Feature Contribution

For training and then testing our approach, we performed a five-fold cross validation with the TUCP-EF corpus to measure the contribution of each of the four subsets of features explained above. Table 3 shows the cross- validation results when cumulatively adding each set of features.

Like other previous contextual classification tasks (e.g., negation detection and assertion classification), using only lexical features provided a good baseline for quantitative assessments classification. However, the lexical features were less helpful for qualitative assessment classification. Surprisingly, using only section features allowed for an F_1-measure of 94% with qualitative assessments. It showed that global features such as section titles could be more beneficial than the local lexical features for qualitative assessments when there are not many distinguishable cue words in the context window surrounding the assessment mention.

Table 3 – Feature contribution: F_1-measure (%)

Features	Quantitative	Qualitative
1. Lexical	90.3	78.3
2. Concept position	57.2	77.2
3. Related concept	78.2	80.2
4. Section	80.9	94.0
1 + 2	91.9	85.9
1 + 2 + 3	91.6	86.9
All (1 + 2 + 3 + 4)	**95.6**	**94.2**

Using concept position features for qualitative assessments performed better than for quantitative assessment. Most summary assessments were mentioned later in the document than other assessments. The features extracted from the nearest LVEF or LVSF mentions were used to mitigate the

limited window size of lexical features. These related concept features helped more than lexical features for qualitative assessments and gave about 2% higher F_1-measures. Combining features, especially with section features, allowed for substantially better performance than with models trained with only individual sets of features. Using a chi-squared test to measure statistical significance, the performance of the full feature system (All, 1 + 2 + 3 + 4) was significantly better than the systems with other feature combinations ($p < 0.001$), but not significantly better than the system using only section for qualitative assessments ($p = 0.801$).

Classification Results

As seen in Table 4, the classification of summary, interpretation, and technical measurement quantitative assessments showed good performance with over 95% F_1-measures. The recommendation category benefitted the most from lexical features with 90% F_1-measure even with few instances (only ten examples). However, the classifier performed very poorly with past finding assessment. No past finding assessment was correctly classified with highly unbalanced class probabilities, indicating that there is ample room for improvement for this category.

Overall, our classifiers achieved good performance, reaching 95.6% F_1-measure for quantitative assessments and 94.2% F_1-measure for qualitative assessments in a five-fold cross-validation evaluation.

Table 4 – Classification results (%): recall (R), precision (P), and F_1-measure (F)

	Quantitative			Qualitative		
Contextual use	**R**	**P**	**F**	**R**	**P**	**F**
Summary	95.9	96.1	96.0	95.7	95.2	95.5
Interpretation	95.8	94.7	95.2	91.7	92.4	92.0
Tech. measurement	97.0	97.3	97.1			
Recommendations	90.0	90.0	90.0			
Past finding	0.0	0.0	0.0			
All	**95.6**	**95.6**	**95.6**	**94.2**	**94.2**	**94.2**

Table 5 displays counts of true positives (bolded), false positives, and false negatives of each category in a confusion matrix. Even though detecting sections played an important role in both quantitative and qualitative assessment classification, some summary assessments under the section titles that were not captured by our section detector were misclassified as interpretation assessments. Summary quantitative assessments were often misclassified as interpretation assessments with 39 false negatives and 32 false positives.

*Table 5 – Confusion matrix [True positives are **bolded**]*

	Classified as				
Contextual use	**Summ**	**Inte**	**Tech**	**Reco**	**Past**
Quantitative					
Summary	**905**	39	0	0	0
Interpretation	32	**888**	7	0	0
Tech. measurement	0	8	**287**	1	0
Recommendations	0	0	1	**9**	0
Past finding	5	3	0	0	**0**
Qualitative					
Summary	**776**	35			
Interpretation	39	**428**			

We observed that many technical measurement assessments were preceded by a colon mark (':') in the corpus. When there is no punctuation between a LVEF and a quantitative assessment, for example in “EF 57%”, the assessment was misclassified as an interpretation assessment. Several past finding assessments were misclassified as summary or interpretation assessments. Future improvement possibilities to reduce these errors include defining features capturing date expressions near the quantitative assessment for this past finding category.

Conclusion

This study demonstrated that the local context of quantitative and qualitative LVEF assessments could be successfully classified using multi-class SVM classifiers. We showed that our application performed well for both quantitative and qualitative assessments with various local and global features. We observed that lexical features and section information features contributed the most to this good performance.

Acknowledgments

This publication is based upon work supported by the Department of Veterans Affairs, HSR&D, grant numbers IBE 09-069 (ADAHF) and HIR 08-374 (Consortium for Healthcare Informatics Research) and HIR 09-007 (Translational Use Case). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or other affiliated institutions.

References

[1] Cheng LT, Zheng J, Savova GK, Erickson BJ. Discerning Tumor Status from Unstructured MRI Reports-Completeness of Information in Existing Reports and Utility of Automated Natural Language Processing. *J Digit Imaging*. 2010; 23(2):119-32.

[2] Nguyen AN, Lawley MJ, Hansen DP, Bowman RV, Clarke BE, Duhig EE, and Colquist S. Symbolic rule-based classification of lung cancer stages from free-text pathology reports. *J Am Med Inform Assoc*. 2010; 17(4): 440-5.

[3] Kim Y, Garvin JH, Heavirland J, Williams J, and Meystre SM. Medication Prescription Status Classification in Clinical Narrative Documents. *Proceedings of AMIA 2014*; 2014.

[4] Mutalik PG, Deshpande A, and Nadkarni PM. Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study Using the UMLS. *J Am Med Inform Assoc*. 2001; 8(6): 598-609.

[5] Chapman WW, Bridewell W, Hanbury P, Cooper GF, and Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*. 2001; 301-10.

[6] Kim J-D, Ohta T, Pyysalo S, Kano Y, and Tsujii J. Overview of BioNLP’09 shared task on event extraction. *Proceedings of the BioNLP 2009 Shared Task*; 2009. p. 1-9.

[7] Farkas R, Vincze V, Móra G, Csirik J, and Szarvas G. The CoNLL-2010 shared task: learning to detect hedges and

their scope in natural language text. *Proceedings of 14th CoNLL*; 2010. p. 1-12.

[8] Vincze V, Szarvas G, Farkas R, Móra G, and Csirik J. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*. 2008; 9(S9).

[9] Kilicoglu H and Bergler S. Syntactic dependency based heuristics for biological event extraction. *Proceedings of the BioNLP 2009 Shared Task*; 2009. p. 119-27.

[10] Morante R, Liekens A, and Daelemans W. Learning the Scope of Negation in Biomedical Texts. *Proceedings of EMNLP 2008*; 2008. p. 715-24.

[11] Uzuner Ö, South BR, Shen S, and DuVall SL. 2010 i2b2/VA Challenge on Concepts, Assertions, and Relations in Clinical Text. *J Am Med Inform Assoc*., 2011;18(5):552-6.

[12] Bruijn BD, Cherry C, Kiritchenko S, Martin J, and Zhu X. Machine learned Solutions for Three Stages of Clinical In- formation Extraction: the State of the Art at i2b2 2010. *J Am Med Inform Assoc*., 2011; 18(5):557-62.

[13] Kim Y, Riloff E, and Meystre SM. Improving classification of medical assertions in clinical notes. *Proceedings of the 49th ACL/HLT*; 2011; 2. p. 311-6.

[14] Bejan CA, Vanderwende L, Xia F, and Yetisgen-Yildiz M. Assertion modeling and its role in clinical phenotype identification. *J Biomed Inform*. 2013; 46(1):68-74.

[15] VistA Monograph.

http://www.ehealth.va.gov/VistA_Monograph.asp

[16] Fan RE, Chang KW, Hsieh CJ, Wang XR, and Lin CJ. LIBLINEAR: A Library for Large Linear Classification. *J Mach Learn Res*, 2008; 9:1871-4.

Address for correspondence

Youngjun Kim,
School of Computing, University of Utah,
50 S. Central Campus Drive, Salt Lake City,
Utah 84112, USA;
youngjun@cs.utah.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-604

Identification of Patients with Family History of Pancreatic Cancer - Investigation of an NLP System Portability

Saeed Mehrabi [a,b], Anand Krishnan[b], Alexandra M Roch[c], Heidi Schmidt[c], DingCheng Li[a], Joe Kesterson[d], Chris Beesley[d], Paul Dexter[d], Max Schmidt[c], Mathew Palakal[b], Hongfang Liu[a]

[a] *Department of Health Sciences Research, Mayo Clinic, Rochester, MN*
[b] *School of Informatics and Computing, Indiana University, Indianapolis, IN*
[c] *Department of Surgery, Indiana University, Indianapolis, IN*
[d] *Regenstrief Institute Inc., Indianapolis, IN*

Abstract

In this study we have developed a rule-based natural language processing (NLP) system to identify patients with family history of pancreatic cancer. The algorithm was developed in a Unstructured Information Management Architecture (UIMA) framework and consisted of section segmentation, relation discovery, and negation detection. The system was evaluated on data from two institutions. The family history identification precision was consistent across the institutions shifting from 88.9% on Indiana University (IU) dataset to 87.8% on Mayo Clinic dataset. Customizing the algorithm on the the Mayo Clinic data, increased its precision to 88.1%. The family member relation discovery achieved precision, recall, and F-measure of 75.3%, 91.6% and 82.6% respectively. Negation detection resulted in precision of 99.1%. The results show that rule-based NLP approaches for specific information extraction tasks are portable across institutions; however customization of the algorithm on the new dataset improves its performance.

Keywords:

Natural language processing, Unstructured Information Management Architecture, Family History, Pancreatic cancer.

Introduction

There has been a slow annual increase in the incidence of pancreatic cancer between 2000-2009 worldwide, in contrast to the decrease for most other major cancers. Pancreatic cancer is one of the deadliest cancers, with approximately 73% death rate among patients within the first year of their diagnosis [1]. It is estimated that 46,420 (23,530 men and 22,890 women) will be diagnosed with, and 39,590 (20,170 men and 19,420 women) will die of, cancer of the pancreas in 2014. Pancreatic cancer has several risk factors such as obesity, smoking, and alcohol intake, but its exact causes are not yet known. Screening the general population for early identification of pancreatic cancer is infeasible, and there is no reliable test for its early detection. Screening high-risk populations might be effective in reducing mortality. It is estimated that 10% of pancreatic cancers have a familial basis [2]. One first-degree relative (parents, siblings or children) with pancreatic cancer increases the risk 7-9 fold, and three or more first-degree relatives with pancreatic cancer increase the risk 17-32-fold [3]. Risk is also increased if a first-degree relative diagnosed with pancreatic cancer before age 50 [4].

Another group of patients that are at risk of having pancreatic cancer is patients with pancreatic cysts [5]. In our previous work, we developed natural language processing (NLP) techniques to identify patients with pancreatic cysts from clinical notes [6-8]. In this study, we are focusing on identifying patients with family history of pancreatic cancer using a rule-based algorithm.

Much of the information in clinical notes is in free text format, making it a challenge for secondary use of clinical data. Information extraction (IE) attempts to structure and encode the information buried in free text clinical notes. Statistical machine learning and rule-based approaches have been used in the development of IE techniques. Machine learning approaches require annotated training examples and lack portability. Rule-based approaches on the other hand perform very well when a task involves a specific subdomain or a limited number of named entities [9]. Although, rule-based approaches are cumbersome to implement, they have been widely used in clinical NLP. In this study, we have developed a rule-based method to identify patients with family history of pancreatic cancer and assessed the generalizability of our algorithm on a different institutions than it was originally developed.

Related Work

Family history identification consists of various steps, including section segmentation, relation discovery between family members and diagnosis, and negation detection. Automatic identification of section headers in clinical notes is an important preprocessing step in the family history extraction. Argumentative zoning is a closely related task that attempts to classify each sentence of a scientific article into one of seven sections of “background”, “other” (other researchers’ work), “own” (author’s work), “aim”, “textual” (textual organization of the paper), “contrast” (work weaknesses of others) and “basis” (authors’ work based on the work of others) [10]. Sequential tagging approaches such as Naïve Bayes (NB) and maximum entropy (MaxEnt) models have been used in solving this problem. MaxEnt model of Merity et al., achieved 96.88% F-Score [11]. Another closely related task is the classification of sentences, in abstracts of scientific articles, into separate sections such as introduction, methods, results, and conclusion. Machine learning algorithms such as SVM, Hidden Markov Models (HMM), and Conditional Random Fields (CRF) have achieved accuracies ranging from 90-94.3% [12-14].

In clinical domains, researchers have developed an algorithm called SecTag that uses a combination of NLP techniques, and rules-based and Naïve Bayesian scoring methods to identify section headers [15]. Section header terminology in this work was developed using the Quick Medical Reference (QMR) knowledge base, Logical Observation Identifiers Names and

Codes (LOINC), and various other resources with data models similar to UMLS [16]. Similar to argumentative zoning, sequential tagging algorithms have also been used in clinical section segmentation. Li et al. used HMM to label sections in clinical notes to one of 15 possible known section types achieving per section accuracy of 93% and per-note accuracy of 70% [17]. Tepper et al. used two methods: A one-step approach that segmented and classified sections in one step, and a two-step approach that used two different models for section segmentation and classification. In the one-step approach, they used the MaxEnt sequential tagging model to identify if a line was in the beginning, inside, or outside (BIO) of a section category. In the two-step approach, they used again MaxEnt sequential tagging to first label each line with BIO tags, and then used a separate classification algorithm to label each section with appropriate section categories. The two-step approach outperformed the one-step approach with precision, recall, F-measure of 90.0-97, 90.4-96.7, 89-96.8 (%) respectively, on three different datasets [18].

Once a family history section is identified and sentences within this section are parsed, the next step is to associate the diagnosis with the correct family members. Both rule-based and dependency parsers have been used to associate family members with diagnoses concepts. Goryachev et al. developed a rule-based algorithm using tokens such as “comma”, “and”, “dot”, “patient has”, “patient had” to assign diagnosis concepts to family members [19]. Their method achieved higher precision and recall in comparison to a dependency parser based algorithm used in another study [20].

Nearly half of the sentences related to family history were negated. Negation detection has been an inevitable step in processing clinical notes that has attracted much attention [21]. NegEx is one of the most commonly used negation algorithms in clinical NLP [22]. Several other negation identification algorithms (such as NegExpander [23], NegFinder [24], ChartIndex [25], DepNeg [26] and DEEPEN [7]) have also been developed using context-free grammar and dependency parsing to improve negation detection accuracy.

To our knowledge, none of the previous work in family history identification consider all of the steps involved in this task. Friedlin et al. reported sensitivity of 93% and positive predictive value of 97% in extraction of family history, but their method only considers family histories reported under the family history section and not those buried under various other sections in clinical notes [27]. It also doesn’t extract exact family members and classify family members as primary, secondary, and unknown relatives. Goryachev et al’s work does not examine the negation status of diagnoses found under the family history section [19]. Lewis et al. reported that only 2% of sentences with a family member term were negated compared to 17% of sentences reported under the family history section being negated. Although they stated their interest to identify negation in their future work, they have not analyzed negation in their latest work [20].

PancPro is a Bayesian modeling framework used to assess the pancreatic cancer risk of patients with family history of pancreatic cancer [28]. However, it does not use NLP techniques to extract family history information from clinical notes, so information was collected using a questionnaire.

Materials and Methods

Data Source

The study was approved by the Institutional Review Board (IRB) protocol of each institution separately. Below are the descriptions of each institution dataset.

Indiana University (IU)

Clinical notes of patients who visited Sidney and Lois Eskenazi Hospital in Indianapolis during March-December 2013 were used in this study. On average, 7,270 patients visited the hospital each month with a range of 80 to 95 thousand reports for all patients during that month. A detailed description of the dataset has been previously published [8]. The dataset was randomly divided into 60% for training and 40% for testing.

Mayo Clinic

We used Mayo cancer registry data to obtain a list of patients with pancreatic cancer. There were a total of 3,573 patients in the registry, out of which 2,923 had a family history section in their clinical notes. Clinical notes for those patients were extracted from the Mayo Clinic data repository, and text from the family history section of those notes form the second data set.

Methods

Clinical reports are organized into sections with headers such as “Physical Examination,” “Medication,” “Family History.” Usually a patient’s family history is reported under the family history section of the narrative reports. However, this is not always the case. It is sometimes mentioned in the patient’s history, diagnosis, or other sections. Based on this understanding, we divided family history identification into two parts: In the first part, the patient’s family history, which is reported under the family history section, was identified. In the second part, the family history section was removed from the clinical note and any mention of a family history in other sections was identified. The first part consisted of three sub-parts: 1) section header detection, 2) family members and diagnoses identification, and 3) relation discovery between family members and diagnoses.

Section header detection

A rule-based algorithm based on the SecTag terminology was developed to identify the clinical notes sections for the IU data set. While at Mayo, clinical notes are CDA 1.0 compliant, wherein sections have been codified. Although SecTag terminology is a large-scale effort to assemble an exhaustive list of terminologies used as section headers of clinical notes, due to lack of standard and universal convention there are still terms used in other institutions that are not found in the SecTag terminology list. For instance, section headers such as “Past Medical, Social, Family History” and “Social and Family History” were used in IU clinical notes, but were not available in the SeTag terminology. We added these terms to our dictionary list to identify family history sections.

Family member and diagnosis identification

After a family history section was identified, sentences reported under this section were detected using Ytex sentence detector [29]. A list of keywords indicating pancreatic cancer concepts and family members were collected using UMLS metathesaurus [30] and manual review of a random set of clinical notes. This dictionary was then used to identify family member and pancreatic cancer concepts within a sentence.

Relation Discovery

Associating family member with pancreatic cancer in a sentence with only one family member is trivial, e.g.:

> A. “Notable for a **father** with what sounds like cirrhosis, colorectal cancer, as well as **pancreatic cancer**, and alcohol abuse.”

However for sentences with more than one family member, this task is challenging (e.g. Sentence B):

B. "The only cancers in her **family** include a first **cousin** on her **mother's** side with breast cancer in her xxx, as well as a **paternal aunt** who had **pancreas cancer** in her xxx, and her **brother** who died of **pancreas cancer** at the age of xxx."

We developed a set of rules that divides the sentence into sub-sentences based on tokens such as ",", ";" or ", and" and associate family member and disease in each sub-sentence.

For example, in sentence "B", after dividing the sentence to three sub-sentences, we could link "paternal aunt" and "pancreas cancer" in the sub-sentence *"as well as a paternal aunt who had pancreas cancer in her xxx",* and the terms "brother" and "pancreas cancer" in the sub-sentence *"and her brother who died of pancreas cancer at the age of xxx"*

If the pancreatic cancer concept was found with no family members in sentences under the family history section, the general term "family history" was assigned to the concept.

In order to identify family history of pancreatic cancer mentioned other in sections of the report, the family history section was removed and the same algorithm was applied where at least one family member must be mentioned.

An NLP system using UIMA framework, shown in Figure 1, was developed to accommodate the above steps. First, two blocks in the UIMA pipeline are 'report separator' and 'metadata annotator' that extract each report's main body and its metadata information, such as report name, ID, date and patient medical record number. Reports' main bodies were then used as an input for the next block of code where family history sections were detected. After the family history section was extracted, the section was split into sentences and family member and diagnosis were identified. We used our previously developed negation algorithm called DEpEndency ParsEr Negation (DEEPEN) to find out the negation status of diagnosis concepts in a sentence [31]. DEEPEN improves the NegEx algorithm by double-checking the negation status of concepts using a nested chain of dependency relations between negation words and desired concepts within a sentence. Finally, all the extracted information (including patient medical number, report name, report date, the sentence containing the concept, the diagnosis concept, and related family members) was found in the sentence, and their negation status was stored in a database.

Figure 1- Analysis engines developed in the UIMA pipeline to identify patients with family history of pancreatic cancer.

Results

Table 1 shows the performance of the system on the IU training and testing sets. The system output consists of patient medical record number, sentence, diagnosis, family member and negation. The results were evaluated as correct or incorrect by two independent reviewers with inter annotator agreement of 95.9%. A result is correct if pancreatic cancer is associated with the correct family member and negation status of the diagnosis was identified accurately. Any errors in these finding were considered as an incorrect instance. We also considered hypothetical cases (i.e. a sister may have had pancreatic cancer.) as incorrect. If pancreatic cancer related to patient or a non-blood relative (e.g., wife or husband) was mentioned, it was considered as irrelevant.

Table 1- IU dataset evaluation.

Train	Correct	Incorrect	Irrelevant	Precision
Affirmed	22	7	2	75.9
Test Set	Correct	InCorrect	Irrelevant	
Affirmed	14	2	2	88.9
Negated	2	0	0	100%

We applied the same algorithm to the Mayo clinic dataset without any modifications (Table 2). Precision is defined as the number of correct instances over the total of correct and incorrect instances. As shown, the performance of the system has been consistent across the two institutions.

Table 2- Mayo dataset evaluation.

	Correct	InCorrect	Irrelevant	Precision
Affirmed	519	72	32	87.8
Negated	438	4	2	99.1

In order to ensure that we did not miss any patient with family history of pancreatic cancer, 100 reports were selected randomly and manually reviewed. Table 3 shows the result of our algorithm to incorporate missing patterns in these reports:

Table 3- Result of Mayo dataset evaluation after system customization.

	Correct	Incorrect	Irrelevant	Precision
Affirmed	550	74	34	88.1
Negated	443	4	2	99.1

Another batch of 100 reports were randomly selected from the Mayo dataset excluding the first 100 reports to manually review the family history of pancreatic cancer. There was no missing pattern in the second set of randomly selected reports.

In relation discovery evaluation, true positives were considered as instances where the pancreatic cancer concept was assigned to the correct family member in the sentence. False negatives were any family member relation that was missed by the system. A wrong family member assignment was considered a false positive.

There were total of 268 patients with a family history of pancreatic cancer out of 3,573 patients with pancreatic cancer in the Mayo Clinic's data set. Figure 2 shows the number of patients identified with first, second, or third degree relative.

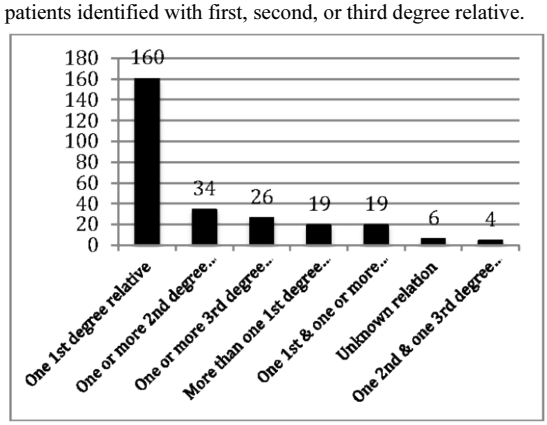


Figure 2- Number of identified patients with one or more 1st, 2nd, or 3rd degree relative.

Table 4- Results of family member identification.

	True Positive	False Positive	False Negative
Family member relation disocvery	579	190	53
	Precision	Recall	F-Measure
	75.3	91.6	82.6

Discussion

In this work, we have developed an NLP system in a UIMA framework with multiple analysis engines, including section segmentation, negation detection, and relation discovery to identify patients with a family history of pancreatic cancer from clinical notes.

We have developed our system on an IU dataset. The IU dataset consisted of any patient who visited the Eskenazi Hospital during 10 months for any reason. Due to low incidence of pancreatic cancer with a familial basis, we had a very low number of patients at IU compared to the Mayo Clinic dataset. Clinical notes at the Mayo Clinic are CDA 1.0 compliant; therefore, section detection developed at IU was not used for the Mayo Clinic dataset. We also did not consider the family history mentions in other sections of clinical notes, other than family history in Mayo Clinic dataset.

We can classify the errors in our system as follows:

1. Sentence Detector

Sentences "C" and "D" show two examples of instances where sentence detector fails to identify the correct boundary of a sentence and therefore the pancreatic cancer concept was improperly assigned to multiple family members.

C) "Father deceased xx-xx
colon polyp pancreatic cancer heart disease alcohol abuse depression
mother
mother alive
heart disease asthma stroke/tia high cholesterol arthritis depression
brothers
2 brothers alive 1 brother deceased
colon polyp asthma
sisters
2 sisters alive
osteoporosis
famhx updates (cvi)
high blood pressure – yes"

D) "His brother died of pancreatic cancer at the age of xx. A sister has a history of breast cancer."

2. Complicated Family Relations

Sentences "*E*", "*F*", and "*G*" show examples of family relation where multiple family member terms were used to show the relation. As we did not have these complicated instances of relationship in our dictionary set, our system related each family member term to the pancreatic cancer separately. For instance, in sentence "E" pancreatic cancer was related to mother, sister, and granddaughter. Sentence "*H*" shows an example where semantic inference is needed to infer that pancreatic cancer is related to the mother.

E) "Recently she found out about her mother's sister's granddaughter who was diagnosed with pancreatic cancer at the age of xx."

F) "He had an uncle that was actually a half-sibling to his mother that died of pancreatic cancer."

G) "He had one cousin on the patient's father's side of the family (the cousin was the son of the patient's father's brother) who had pancreas cancer at age xx."

H) "Her son (our patient) found her deceased about xx p.m. a postmortem examination showed cause of death was due to multiple blood clots and she was found to have a widespread pancreatic cancer."

3) System Failure

As mentioned in the relation discovery section, a set of rules was developed to divide the sentence into sub-sentences. When there are multiple family relation terms in a sentence such as sentence "I". Each family relation term was then associated with the pancreatic cancer concept within the sub-sentence. In sentence "I," "*pancreatic cancer*" is associated with "*paternal grandfather,*" but it failed to associate the "mother" and "father" with the pancreatic cancer concept in the sub-sentence *"his mother, father,"* because there is no pancreatic cancer concept in the sub-sentence.

I) "His mother, father, and paternal grandfather died from pancreatic cancer."

There were few instances where co-referencing was needed to extract the right family relation (see sentence "J"). Our current system does not handle co-referencing.

J) "She has one son living and one deceased. the one that is living has a recent diagnosis of pancreatic cancer, and three daughters."

Conclusions

Pancreatic cancer is referred to as silent killer due to its few sign and symptoms until it is in well-advanced cancer stages. Screening the general population for pancreatic cancer is not feasible because of its low incidence and the lack of effective screening tests. Pancreatic cyst and family history of pancreatic cancer represent two windows of opportunity for early detection of pancreatic cancer. We have developed a rule-based algorithm to identify patients with a family history of pancreatic cancer retrospectively from their clinical records. Development of clinical NLP system requires resources, such as domain experts to develop guidelines, nurse abstractors to create gold standards, and researchers/programmers to develop and analyse the system. Although rule-based methods highly depend on the natural language that they have been developed on, this study shows that as long as the rules are kept simple and generalizable, we can transfer an algorithm developed in one institution to other institutions.

Future steps involve refinement of the family relation discovery rules, especially regarding the sentence detection algorithm. A risk stratification method will also be developed based on the number and degree of family relations to assess patients' risk of having cancer and a surveillance strategy will be designed to follow up with patients according to their risk.

Acknowledgments

This work was made possible by funding from NIH R01GM102282, R01LM11369, R01LM11829, and R01 LM011934.

References

[1] American Cancer Society. Cancer Facts & Figures American Cancer Society, Atlanta, 2014.

[2] Permuth-Wey J and Egan KM. Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. Fam Cancer 2009;8(2):109-17.

[3] Shi C , Hruban RH, and Klein AP. Familial pancreatic cancer. Arch Pathol Lab Med 2009: 133(3): 365-74.

[4] Klein AP et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res 2004: 64(7): 2634-8.

[5] Pandol S et al. Epidemiology, risk factors, and the promotion of pancreatic cancer: role of the stellate cell. J Gastroenterol Hepatol 2012: 27: 127–134.

[6] Al-Haddad MA et al. Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. International Hepato-Pancreato-Biliary Association 2010: 12(10): 688-95.

[7] Mehrabi S et al. An efficient pancreatic cyst identification methodology using natural language processing. Stud Health Technol Inform 2013: 192: 822-6.

[8] Roch AM et al. Automated pancreatic cyst screening using natural language processing: a new tool in pancreatic cancer early detection. HPB doi: 10.1111/hpb.12375.

[9] Liu H et al. An information extraction framework for cohort identification using electronic health records. AMIA Jt Summits Transl Sci Proc 2013:149–153.

[10] Teufel S and Moens M. Summarizing scientific articles: experiments with relevance and rhetorical status. Computational Linguistics 2002: 28(4): 409-445.

[11] Merity S, Murphy T, and Curran JR. Accurate Argumentative Zoning with Maximum Entropy models. Text and Citation Analysis for Scholarly Digital Libraries (NLPIR4DL), 2009.

[12] Lin J, Karakos D, Demner-Fushman D, and Khudanpur S. Generative content models for structural analysis of medical abstracts. HLT-NAACL BioNLP Workshop. 2006:65–72.

[13] Hirohata KK, Okazaki N, Ananiadou S, and Ishizuka M. Identifying sections in scientific abstracts using conditional random fields. International Joint Conference on Natural Language Processing 2008: 381–388.

[14] McKnight L and Srinivasan P. Categorization of sentence types in medical abstracts. AMIA Annu Symp Proc. 2003:440–444.

[15] Denny JC et al. Evaluation of a method to identify and categorize section headers in clinical documents. JAMIA 2009;16(6):806-15.

[16] Denny JC, Miller RA, Johnson KB, and Spickard A. Development and evaluation of a clinical note section header terminology. AMIA Annu Symp Proc. 2008:156-60.

[17] Li Y, Gorman SL, Elhadad N. Section classification in clinical notes using supervised hidden markov model. ACM International Health Informatics Symposium. 2010: VOL???: 744–750.

[18] Tepper M, Capurro D, Xia F, Vanderwende L, and Yetisgen-Yildiz M. Statistical Section Segmentation in Free-Text Clinical Records. Eight International Conference on Language Resources and Evaluation. 2012:2001-2008.

[19] Sergey Goryachev, Hyeoneui Kim, and Qing Zeng-Treitler. Identification and Extraction of Family History Information from Clinical Reports. AMIA Annu Symp Proc. 2008:247–251.

[20] Lewis N, Gruhl D, and Yang H. Dependency Parsing for Extracting Family History. IEEE HISB 2011: VOL????: 237-242.

[21] Wu S et al. Negation's Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing. PLoS ONE 2014: 9(11): pages?????.

[22] Chapman WW, Bridewell W, Hanbury P, Cooper GF, and Buchanan BG. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. Journal of Biomedical Informatics 2001: 34(5): 301–310.

[23] Aronow DB, Fangfang F, and Croft WB. Ad Hoc Classification of Radiology Reports. J Am Med Inform Assoc 1999: 6(5) Sep-Oct: 393–411.

[24] Mutalik PG, Deshpande A, and Nadkarni P. Use of general purpose negation detection to augment concept indexing of medical documents: a quantitative study using the umls. J Am Med Inform Assoc 2001: 8: 589-609.

[25] Huang Y and Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. J Am Med Inform Assoc 2007: 14: 304-311.

[26] Sunghwan S, Stephen W, and Chute CG. Dependency Parser-based Negation Detection in Clinical Narratives. AMIA Summits Transl Sci Proc. 2012:1–8.

[27] Friedlin J and McDonald CJ. Using a natural language processing system to extract and code family history data from admission reports. AMIA Annu Symp Proc. 2006:925.

[28] Wang W et al. PancPRO: risk assessment for individuals with a family history of pancreatic cancer. J Clin Oncol 2007: 25(11): 1417-22.

[29] Garla V et al. The Yale cTAKES extensions for document classification: architecture and application. J Am Med Inform Assoc 2011: 18(5) Sep-Oct: 614-20.

[30] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004: 32 Jan: 267-70.

[31] Mehrabi S et al. DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx JBI 2015 doi:10.1016/j.jbi.2015.02.010.

[32] M C de Marneffe , B MacCartney , and C D Manning , "Generating Typed Dependency Parses from Phrase Structure Parses," in LREC, 2006.

[33] Robert J Carroll et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc 2012: 19: e162-169.

Address for correspondence

Mayo Clinic, 200 1st SW, Rochester MN 55905, Tel: 507-773-0057
Email address: liu.hongfang@mayo.edu (Dr. Hongfang Liu).

535 W. Michigan St, Indianapolis, IN, 46202, USA

Tel: 317-278-7689, Fax: 317-278-7669 (Dr. Mathew Palakal)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-609

Heart Failure Medications Detection and Prescription Status Classification in Clinical Narrative Documents

Stéphane M. Meystre, MD, PhD [a,c], Youngjun Kim, MS [b,c], Julia Heavirland [c], Jenifer Williams [c], Bruce E. Bray, MD [a,d], Jennifer Garvin, PhD, MBA [a,c]

[a] *Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA*
[b] *School of Computing, University of Utah, Salt Lake City, Utah, USA*
[c] *VA Health Care System, Salt Lake City, Utah, USA*
[d] *Cardiovascular Medicine, School of Medicine, University of Utah, Salt Lake City, Utah, USA*

Abstract

Angiotensin Converting Enzyme Inhibitors (ACEI) and Angiotensin II Receptor Blockers (ARB) are two common medication classes used for heart failure treatment. The ADAHF (Automated Data Acquisition for Heart Failure) project aimed at automatically extracting heart failure treatment performance metrics from clinical narrative documents, and these medications are an important component of the performance metrics. We developed two different systems to detect these medications, rule-based and machine learning-based. The rule-based system used dictionary lookups with fuzzy string searching and showed successful performance even if our corpus contains various misspelled medications. The machine learning-based system uses lexical and morphological features and produced similar results. The best performance was achieved when combining the two methods, reaching 99.3% recall and 98.8% precision. To determine the prescription status of each medication (i.e., active, discontinued, or negative), we implemented a SVM classifier with lexical features and achieved good performance, reaching 95.49% accuracy, in a five-fold cross-validation evaluation.

Keywords:

Heart Failure; Angiotensin Converting Enzyme Inhibitors; Angiotensin II Receptor Blockers, Biomedical Informatics; Natural Language Processing.

Introduction

Heart Failure (HF) is characterized by the inability of the heart to pump blood at a rate sufficient to answer metabolizing tissues' needs. It is a frequent condition in the U.S. adult population, causing more hospitalizations than all forms of cancer combined [1]. HF treatment can include dietary and physical activity therapies and invasive therapies, but pharmacologic therapies are the most common. Among pharmacologic therapies, Angiotensin Converting Enzyme Inhibitors (ACEI) are a mainstay of treatment in patients who can tolerate them; for patients who cannot take these drugs, Angiotensin Receptor Blockers (ARB) agents offer an alternative. Assessment of the care HF patient's benefit from relies on detecting evidence of current treatment with these medications in clinical notes, along with analysis of the status of these medications. This information, combined with the functional assessment of the left ventricular function, is a key component of HF treatment performance measures.

This study was realized in the context of the ADAHF (Automated Data Acquisition for Heart Failure) project, a U.S. Department of Veterans Administration (VA) project that aims to automatically extract data for HF treatment performance measures from clinical notes. These performance measures include left ventricular ejection fraction (LVEF) assessments (their mention and measured values), medications (ACEIs and ARBs), or reasons not to administer these medications, as part of the Joint Commission National Hospital Quality Measures Heart Failure Core Measure Set the VA has adopted for use in Veteran care.

Previous efforts have focused on detecting LVEF mentions and values in clinical notes and determining their abnormality (LVEF < 40%) [2-4].

However, to assess whether each patient's clinical record substantiates proper treatment for HF, active medications should be considered as well. In this study, we focused on ACEIs and ARBs detection and classification of the status of these detected medications.

Clinical notes are mostly composed of unstructured text and have multiple different formats depending on specialties and institutions. Misspelled words and non-narrative text formats can be found in clinical notes, especially when directly typed into the system by healthcare providers. For example, one of the medications included in our study, "Losartan", was found with the following issues:

"Lorsartan" (misspelled word)

"Losartan30 mg" (tokenization problem: missing whitespace)

"L [newline] osartan" (word wrapping problem: word wrongly split)

These problems make medication detection based on a simple dictionary lookup more difficult.

Medications can be mentioned in multiple contexts in clinical notes. Most are currently taken by the patient (i.e., active), but some can be mentioned in the patient medical history as discontinued, or even mentioned as specifically not taken by the patient, because of an allergic reaction for example. This contextual information, the status of the medication, therefore, needs accurate analysis, and we designated three different categories: active (the patient currently takes the medication),

discontinued (the patient remains off the medication or is temporarily taken off the medication), and negative (the medication does not pertain to the patient or is negated).

The automatic extraction of medication information from clinical notes was the main task of the 2009 i2b2 NLP challenge [5]. It focused on the identification of medications and attributes including dosage, frequency, treatment duration, mode of administration, and the reason for the administration of the medication. Almost twenty teams participated in this challenge, and Meystre and colleagues built a system called Textractor that combined dictionary lookups with machine learning approaches and reached a performance of about 72% recall and 83% precision [6].

Patrick and Li trained a sequence-tagging model using conditional random fields (CRF) to detect medications with various lexical, morphological, and gazetteer features. Their tagger reached about 86% recall and 91% precision (the highest score in the challenge) [7].

No published research attempted prescription status classification, but some developed systems to recognize the context or assertions of medical concepts. Chapman and colleagues created the NegEx algorithm, a simple rule-based system that uses regular expressions with trigger terms to determine whether a medical term is negated. They reported 77.8% recall and 84.5% precision for medical problems in discharge summaries [8]. Chapman and colleagues also introduced the ConText algorithm, which extended the NegEx algorithm to detect four context categories: negated, hypothetical, historical, and not associated with the patient [9].

Kim and colleagues [10] developed a Support Vector Machines (SVM [11])-based assertion classifier for the 2010 i2b2 NLP challenge [12] and their system reached 94.17% accuracy by regulating un-balanced class probabilities and adding features to improve performance recognizing minority classes. Our system for medication status classification expanded this system with a simplified feature set.

In the following sections, we will describe the methods we used for medication detection and prescription status classification and present our experimental results.

Methods

Materials

The ADAHF project included the development of a large annotated corpus of clinical narrative text notes from patients with HF treated in a group of VHA medical centers in 2008. Each document in this corpus was annotated by two reviewers independently, and a third reviewer adjudicated disagreements.

We randomly sampled 3,000 clinical notes from our training corpus. The most common clinical note types were progress notes, discharge summaries, history and physical notes, cardiology consultation notes, and echocardiogram reports. These 3,000 notes included 6,007 medication annotations (4,911 ACEIs and 1,096 ARBs). Medications annotated in our project included all ACEI and ARB preparations available in the U.S. The distribution of the most common medications annotated in our corpus is presented in Table 1.

Each annotated medication was also assigned a status category: active, discontinued, or negative. Among annotated medications, 4,491 (74.76%) were active, 1,191 (19.83%) were discontinued, and 325 (5.41%) were negative. Even though the discontinued or negative status was not common, they have to be classified accurately.

Table 1 – Most Frequent Medications with Term Variants

Medication	Type	%	Frequent term variants
Lisinopril	ACEI	39.9	Lisinopril
ACEI	ACEI	16.7	ACE, ACEi, ACE inhibitor
Benazepril	ACEI	16.2	Benazepril, Lotensin
Losartan	ARB	8.2	Losartan
Fosinopril	ACEI	6.0	Fosinopril, Fos
Valsartan	ARB	5.0	Valsartan, Diovan
ARB	ARB	4.5	ARB, Angiotensin receptor blocker
Captopril	ACEI	1.6	Captopril
Enalapril	ACEI	0.8	Enalapril
Irbesartan	ARB	0.4	Irbesartan
Others		0.7	

Rule-Based Medication Detection: Our first approach was based on a dictionary lookup with pre-defined medication entries. We used a modified version of ConceptMapper,[13] a highly configurable UIMA [14] dictionary annotator.

To build this baseline system (**DL1 system**), the dictionary of medication terms had to be manually built. For each ACEI and ARB medication, we included generic names, brand names, and other frequently used name variations (from RxNorm and clinical experts' experience with clinical text).

Misspelled medication names are common in our corpus. For example, "Lisinopril", the most frequent medication had 21 different misspellings in our corpus:

Lisinipril	Lisinoppril	Lisnopril
Lasinopril	Lisiniprli	Lisinopiril
Lisinorpril	Linsinopril	Liinopril
Linsopril	Lisiniopril	Lisiniprol
Lisinoril	Lisinorpil	Lisinorpirl
Lisinpril	Lisionpril	Lisniopril
Lisnoril	Lisonopril	Loisinopril

To improve the sensitivity of our detection and include misspellings, we added fuzzy string searching for spelling variant replacement (**DL1 + fuzzy searching system**). We used the edit distance (or Levenshtein distance)[15], the minimum number of single-character edits needed to transform one word into another, to check whether each text token was a spelling variant of our pre-defined medication terms.

To reduce the number of false positive matches in a second version of our system (**DL2 system**), we only considered matched word tokens as medication name candidates when they met the following criteria: 1) the first character was matched or the last four characters matched one of our dictionary terms, and 2) the edit distance between the word token and one of our dictionary terms was less than two.

In addition to fuzzy searching for each word token, we also analyzed all tokens separated by a newline to reconstruct wrongly split words like "L [newline] osartan." After removal of the newline character between two word tokens, treating them as a one word token, we considered them as a correctly reconstructed word if they met the two-step criteria specified above.

These fuzzy searching strategies allowed detection of more medications even when they were misspelled or erroneously split by a newline character.

Machine Learning-Based Medication Detection: The second approach for medication detection was based on machine learning methods. We built a token-based classifier without sequential learning. A sequential tagger requires sentences as inputs and predicts the sequence of labeled tags with probabilities. In a Hidden Markov Model [16], a popular choice for sequence tagging, transition probabilities from one tag to the next tag are considered when learning a model with a dynamic learning algorithm like Viterbi path [17].

In token-based learning, the tagger only predicts a label for each token, independently from the previous tags. We used LIBLINEAR[18] with a linear SVM classifier to train our token-based model (**SVM system**). We used lexical features (the word itself, two words preceding it, and two words following it) and morphological features (prefix and suffix up to length of five) with B-I-O tags (B denotes the beginning of a term, I a token inside a term, and O a token outside a term). Because the classifier predicts each tag independently from the previous tags, we did some post-processing (using a few heuristic rules) to avoid undesirable tag sequences like a "B-ACEI" (token at the beginning of an ACEI name) followed by "I-ARB" (token inside an ARB name).

Finally, we combined the rule-based and machine learning-based methods by using DL2 system's predictions as new features for the SVM classifier (**SVM + DL2 system**). The feature vector was augmented with the DL-II system predictions for the current token, the two previous tokens, and the two following tokens.

Medication Status Classification: To classify medications as active, discontinued, or negative (details in the Introduction), pre-processing included tokenization with a modified version of the cTAKES [19] tokenizer. The prescription status classifier only used lexical features from the tokenizer (i.e., all word tokens). We also used LIBLINEAR for this task, with a wider context window than for medication detection (five preceding and five following words) and no morphological features.

Metrics and Statistical Analysis: Accuracy of the detection of medications is reported using typical metrics for information extraction or retrieval: *Recall* (equivalent to sensitivity in this context; equals true positives/(true positives + false negatives)), *Precision* (equivalent to positive predictive value on this context; equals true positives/(true positives + false positives), and the *F_1-measure* (harmonic mean of recall and precision; equals (2*recall*precision)/(recall+precision) when giving equal weight to recall and precision). These metrics were *macro-averaged* to obtain average values for each system (i.e., each metric was calculated for each document, and then averaged across all 3,000 documents).

Descriptive statistics are reported with 95% confidence intervals. Statistical analysis to compare our different approaches to detect medications was realized using the Student's t-test as well as the Mann-Whitney U test for its higher efficiency with non-normal distributions.

Results

Medication Detection: As an easily accessible baseline system for our evaluation, we used eHOST[20], the Extensible Human Oracle Suite of Tools, an open source text annotation tool, to detect medications with a pre-compiled dictionary of medication terms, as specified in our annotation guideline.

Table 2 – Five-fold Cross Validation Results for Medication Detection (macro-averaged percentages)

System	R	P	F
eHOST	83.7 ±1.1	87.2 ±1.1	84.8 ±1.1
DL1	94.8 ±0.7	96.7 ±0.6	95.3 ±0.6
DL1+fuzzy searching	96.6 ±0.5	97.0 ±0.5	96.5 ±0.5
DL2	99.2 ±0.3	98.6 ±0.3	98.7 ±0.3
SVM	97.9 ±0.4	97.4 ±0.4	97.5 ±0.4
SVM+DL2	99.3 ±0.2	98.8 ±0.3	98.9 ±0.2

R=Recall, P=Precision, F=F_1-measure. Percentages reported with 95% confidence intervals

This dictionary listed multiple terms for 44 different medications and general categories. eHOST reached moderate performance (Table 2, Figures 1 and 2).

However, many medications detected by eHOST would be considered true positives if partial span matches were counted. For example, medication names are sometimes attached to punctuation in our corpus (e.g., "Losartan;", "Losartan)") and punctuation (; or) in the same examples) should be excluded for exact matches, but eHOST detected these medications with the punctuation characters.

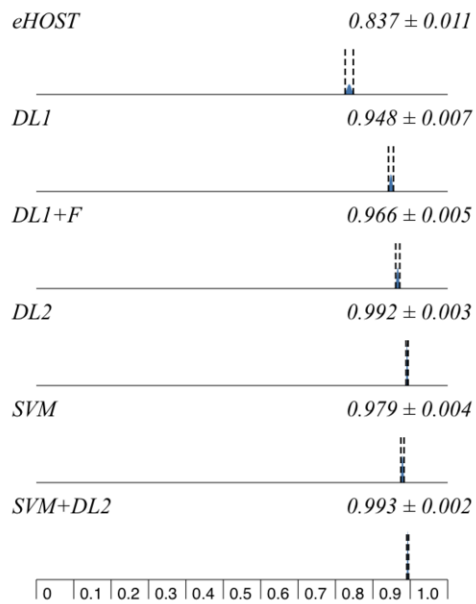


Figure 1 - Systems Recall Comparison

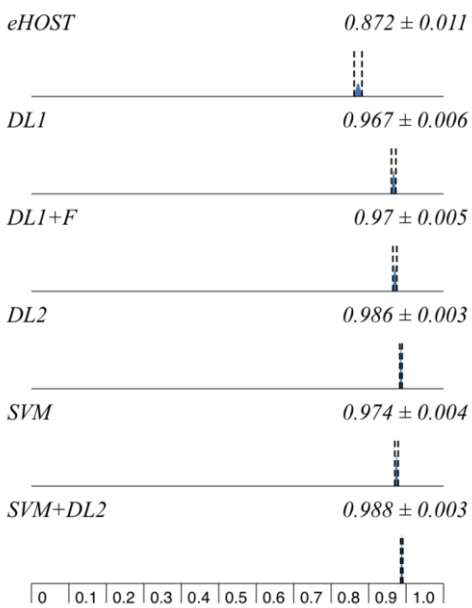


Figure 2 - Systems Precision Comparison

Our first dictionary lookup system (DL1) reached good performance, especially for precision (96.7%). Adding fuzzy string searching to DL1 increased recall to 96.6%. When adding wrongly split words correction (DL2), recall reached more than 99% by detecting medication names that contained newline characters.

When evaluating the machine learning-based system with a five-fold cross validation, the SVM classifier achieved lower recall and precision than the DL2 system, but closer to the DL1 system with fuzzy searching. SVM with DL2 output allowed for the best performance, with the highest recall at 99.3% and a 98.9% F_1-measure.

Table 3 – Systems Recall Pairwise Comparison

Recall	eHOST	DL1	DL1+F	DL2	SVM
DL1	*				
DL1+F	*	*			
DL2	*	*	*		
SVM	*	*	*	*	
SVM+DL2	*	*	*	0.368	*

Precision	eHOST	DL1	DL1+F	DL2	SVM
DL1	*				
DL1+F	*	0.411			
DL2	*	*	*		
SVM	*	0.047	0.233	*	
SVM+DL2	*	*	*	0.306	*

* <0.001

As indicated in Table 3, pairwise comparison and statistical analysis of the results reported in Table 2 and Figures 1 and 2 demonstrated that all differences were significant ($p<0.001$) except all metrics between DL2 and SVM+DL2 (p=0.306-368), and precision between DL1 and DL1+fuzzy searching (p=0.411), and between DL1+fuzzy searching and SVM (p=0.233).

Medication Status Classification: We also used a five-fold cross validation with the 6,007 medication annotations to measure performance of medication status classification (Table 4). The overall accuracy was 95.49%. Precision of each status was above 90%, and recall of the *discontinued* status was 86.23%. Interestingly, recall was higher than precision with the *negative* status, even though they were associated with only 5.41% of the annotated medications in our corpus.

Table 4 – Five-fold Cross Validation Results for Medication Status Classification

Medication Status	**R**	**P**	**F**
Active	98.0	96.3	97.1
Discontinued	86.2	92.9	89.5
Negative	94.2	93.3	93.7
Overall	95.5	95.5	95.5

R=Recall, P=Precision, $F=F_1$-measure.

Discussion

The dictionary lookup approach for medications detection was very efficient, and the fuzzy string searching boosted performance, especially recall. The main advantages of this method is that there is no need for manually annotated training examples and processing is fast. With the SVM classifier, morphological features helped detect misspelled medication names but didn't help with erroneously split medication names. We observed that a machine learning-based method could be applied successfully to this task without an external medical knowledge base. But it didn't add any significant performance improvement when compared to the improved dictionary-based and rule-based system (DL2). Overall, it and was probably not a worthy effort in our case, considering the requirement for annotated training examples to train the SVM classifier.

Medication status classification was satisfactory, and even though performance with *active* and *negative* cases was quite good, there is ample room for improvement with the *discontinued* status. A total of 230 (71+159) *active* or *discontinued* cases were misclassified as the other class (Table 5).

Table 5 –Medication Status Classification Confusion Matrix

	Classified as		
	Active	**Discontinued**	**Negative**
Active	4403	71	17
Discontinued	159	1027	5
Negative	12	7	306

One possible avenue for future work is to develop specific patterns or lexicons for this *discontinued* status, including terms like 'hold', 'discontinue', or 'd/c'. Recognizing clinical document sections or detecting phrases mentioning why the

patient was not on the medication might play an important role as the classifier.

Our experimentation with machine learning-based approaches to detect specific medications was limited to one method, SVMs. Other machine learning algorithms such as Conditional Random Fields have been successfully applied to similar tasks and could also be applied to detect ACEIs and ARBs.

Conclusion

This study showed that information extraction methods using rule-based or machine learning-based approaches could be successfully applied to the detection of ACEI and ARB medications in unstructured and somewhat messy clinical notes. We boosted medication detection performance with fuzzy string searching and combining the two approaches. The preliminary work to classify the status of each medication showed that the words surrounding medication names were the most beneficial features.

Acknowledgments

This publication is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, HSR&D, grant numbers HSR&D IBE 09-069. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the University of Utah School of Medicine.

References

[1] Chen J, Normand S-LT, Wang Y, Krumholz HM. National and regional trends in heart failure hospitalization and mortality rates for Medicare beneficiaries, 1998-2008. JAMA 2011; 306:1669–1678.

[2] Garvin JH, DuVall SL, South BR, Bray BE, Bolton D, Heavirland J, Pickard S, Heidenreich P, Shen S, Weir C, Samore M, Goldstein MK. Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. J Am Med Inform Assoc 2012; 19:859–866.

[3] Meystre SM, Kim J, Garvin J. Comparing Methods for Left Ventricular Ejection Fraction Clinical Information Extraction. AMIA Summits Transl Sci Proc, CRI 2012:138.

[4] Kim Y, Garvin JH, Heavirland J, Meystre SM. Relatedness Analysis of LVEF Qualitative Assessments and Quantitative Values. AMIA Summits Transl Sci Proc, CRI 2013:123.

[5] Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010; 17:514–518.

[6] Meystre SM, Thibault J, Shen S, Hurdle JF, South BR. Textractor: a hybrid system for medications and reason for their prescription extraction from clinical text documents. J Am Med Inform Assoc 2010; 17:559–562.

[7] Patrick J, Li M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. J Am Med Inform Assoc 2010; 17:524–527.

[8] Chapman W, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001; 34:301–310.

[9] Harkema H, Dowling JN, Thornblade T, Chapman W. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform 2009;42:839-851.

[10] Kim Y, Riloff E, Meystre SM. Improving classification of medical assertions in clinical notes. HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011: 311-316.

[11] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. New York, New York, USA: ACM Press; 1992:144–152.

[12] Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011; 18:552–556.

[13]ConceptMapper. Available at: http://uima.apache.org/d/uima-addons-current/ConceptMapper/ConceptMapperAnnotatorUserGuide.html

[14] Apache UIMA. Available at http://uima.apache.org

[15] Levenshtein VI. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Soviet Physics Doklady 1966; 10:707.

[16] Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics 1966.

[17] Viterbi AJ. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. Information Theory, IEEE Transactions on 1967; 13:260–269.

[18] Chang C-C, Lin C-J. LIBSVM : a library for support vector machines; 2001. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[19] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010; 17:507–513.

[20] South BR, Shen S, Leng J, Forbush TB, DuVall SL. A Prototype Tool Set to Support Machine-Assisted Annotation. BioNLP '12 Proceedings of the 2012 Workshop on Biomedical Natural Language Processing 2012;130–139.

Address for correspondence

Stephane Meystre, University of Utah, Biomedical Informatics Department, 421 Wakara Way, Suite 140, Salt Lake City, UT 84108, USA

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-614

Classifying the Indication for Colonoscopy Procedures: A Comparison of NLP Approaches in a Diverse National Healthcare System

Olga V. Patterson, PhD[a,b], Tyler B. Forbush, RCP[a], Sameer D. Saini, MD, MS[c,d], Stephanie E. Moser[d], Scott L. DuVall, PhD[a,b]

[a]VA Salt Lake City Health Care System; [b]University of Utah, Salt Lake City, UT;
[c]VA Ann Arbor HSR&D Center for Clinical Management Research; [d]University of Michigan Medical School, Ann Arbor, MI

Abstract

In order to measure the level of utilization of colonoscopy procedures, identifying the primary indication for the procedure is required. Colonoscopies may be utilized not only for screening, but also for diagnostic or therapeutic purposes. To determine whether a colonoscopy was performed for screening, we created a natural language processing system to identify colonoscopy reports in the electronic medical record system and extract indications for the procedure. A rule-based model and three machine-learning models were created using 2,000 manually annotated clinical notes of patients cared for in the Department of Veterans Affairs. Performance of the models was measured and compared. Analysis of the models on a test set of 1,000 documents indicates that the rule-based system performance stays fairly constant as evaluated on training and testing sets. However, the machine learning model without feature selection showed significant decrease in performance.

Therefore, rule-based classification system appears to be more robust than a machine-learning system in cases when no feature selection is performed.

Keywords:

Natural Language Processing; Colonoscopy; Screening.

Introduction

Colonoscopy is a common procedure that aims to detect abnormalities of the colon through direct visual inspection [1]. One of the most common and important reasons to perform colonoscopy is to screen for colorectal cancer [2]. While regular colorectal cancer screening is widely recommended, recent studies have also raised concerns about overuse of screening colonoscopy [3]. As with any clinical procedure, colonoscopy carries risks of both minor and major complications, including death [4,5].

The appropriate timing of colonoscopy depends on the reason or indication for the colonoscopy. Previous studies of colonoscopy overuse have relied on administrative data to determine whether a colonoscopy was performed for a screening or non-screening indication [6]. But studies have repeatedly demonstrated that relying strictly on administrative codes achieves a relatively low accuracy for screening indication (sensitivity of 70.1% and specificity of 71.6%) [7,8]. And even when additional diagnostic codes were included in the algorithms, sensitivity went up only slightly to 83% and specificity to 76% in the Department of Veterans Affairs (VA) clinical setting [9].

The Quality Assurance Task Group of the National Colorectal Cancer Roundtable recommends a list of key areas that should be covered in colonoscopy procedure reports, one of which is a detailed description of procedure indication [10]. Recording the data in structured format is not always supported by clinical information systems to the extent needed for proper documentation of patient risk factors. Thus, these clinical reports are commonly recorded in narrative format and stored as unstructured text.

Background

Data Sources

As the largest integrated healthcare network in the United States, VA has health information on millions of Veterans for which it provides care. A benefit of the integrated healthcare system is seen in the comprehensive electronic medical record, the Veterans Health Information Systems and Technology Architecture (VistA), which was established in 1990's and is used across all VA healthcare facilities. Within the VA healthcare system, there are more than 1,400 hospitals, clinics, and nursing homes nationwide. Since US Veterans tend to receive a large proportion of their care within the VA system, the patient medical records contain a detailed history of care over the last 20 years. The records are aggregated into the VA Corporate Data Warehouse, which contains patient records from all VA facilities from October 1999 to present. These records are made available to VA-credientialed researchers with study approvals through the VA Informatics and Computing Infrastructure (VINCI), an analytic environment that combines access to nationwide VA patient records with computational resources and software tools.

Even though common infrastructure and unique patient identifier enable integration of all data at the database level, individual sites have the flexibility to modify data entry in VistA through templates and forms [11]. Because of the strict requirement to comply with interoperability requirements for structured data, the variability among different sites mostly affects free text entries [12]. The divergence of installations across all sites and the large number of clinicians across all facilities result in inconsistent format and large amount of information recorded in clinical notes in the system [13]. This same variation exists in the way documents are named – making it difficult to identify relevant documents.

Colonoscopy indication is typically recorded in procedure reports in narrative format and is available for extraction with natural language processing (NLP). Document level classification is an NLP task that utilizes the information contained in a document to assign a label from two or more labels to each document. Since each colonoscopy report presents information about a single colonoscopy procedure, determining the indication for the procedure as recorded by

clinicians can be treated as a classification task. Each document can be classified as containing evidence of the purpose for the procedure as being average-risk screening ("Screening") or high-risk screening, surveillance, or diagnosis ("Non-screening"). Detecting relevant documents is hindered by data entry variability and the lack of standard document titles, so another classification label, "Non-colonoscopy", was added to filter out documents that are not colonoscopy reports. A final classification label of "Unknown" was given when the indication of the colonoscopy was not given.

Classification algorithms

Detecting colonoscopy overuse at a population level and tracking the practice changes over time requires an automated system to perform extraction of colonoscopy indication from a heterogeneous dataset. A number of approaches have been used for document classification, including rules and a variety of supervised machine-learning algorithms [14]. Each of these approaches has its advantages and limitations.

The heuristic rule-based approach (also called knowledge engineering) is one of the most established computational methods for document classification [15]. This approach utilizes expertise of language engineers to design lexicons and classification logic in a form of "if-then" decision tree specifically for the target domain. The advantages of such an approach are a) the classification logic is intuitive and understandable for humans, b) the algorithm is straightforward to implement, and c) model development does not require a large annotated corpus [16.17]. It has been suggested that rule-based approaches can be applied on a new corpus without changes, because they rely on grammatical and lexical structures inherent to the specific natural language and not on distribution of specific features in the target corpus [18]. The main limitation of this approach is that the initial rule development is time-consuming and effort-intensive in case of a broad domain and highly variable language. However, the language used by clinicians is considered to be a sublanguage, and especially in such focused subdomain as colonoscopy, it can be expected to be lexically restricted [19].

Supervised machine learning approaches create a statistical model using computational methods applied on manually annotated corpora. The main benefit of this approach is that the resulting model is grounded in the objective mathematical representation of the target language. Availablitly of open source implementations of a number of algorithms in a variety of programming languages simplifies development of classification systems [20,21]. Minimization of the human effort requried in knowledge base development is often mentioned as a benefit of using a machine learning approach [15]. However, such a statement misrepresents the actual human effort required for successful implementation of a classification system. In order for the statistical algorithm to have a sufficient number of training examples, a relatively large number of documents have to be manually annotated. While generally speaking, the annotators' time is less expensive than the time of a highly qualified clinical expert and knowledge engineer, the need for a large, consistently annotated corpus puts a high time and effort demand on annotators for planning, training, and adjudication tasks. At the same time, machine learning approaches do not eliminate knowledge engineering burden, but rather shift it to feature engineering [14]. The easiest and the most popular feature set to implement is known as a bag-of-words model that simply uses the lexical shape of all tokens in a document as binary features [22]. Such models result in a large and very sparse feature space because each distinct token in the whole corpus becomes a feature and each document has only a subset of the lexicon. However, in order to make classification more efficient and accurate, thoughtful feature selection is required [23]. Depending on the feature set, additional annotations might be needed, increasing the human effort [24]. Another drawback of machine learning approaches is that typically a model is highly dependent on the distribution of the classes in the training dataset, therefore, adapting the model for a dataset with a different distribution will likely lead to a decreased performance. From the end user perspective, the machine learning models are hard to interpret and appear to be "black boxes" that can be mistrusted by clinicians [25].

The Automated Retrieval Console (ARC) is an information retrieval system developed in VA that allows binary document classification by applying a supervised machine-learning algorithm on a set of features derived through NLP [26]. ARC utilizes the Apache Clinical Text Analysis and Knowledge Extraction System (cTAKES) to define features without the need for customization [27]. ARC evaluates feature combinations and allows the user to select the highest performing feature set or any of the other feature set combinations for the classification model. By packaging NLP and machine learning components into a unified application, ARC shifts users' development focus to document selection and reference standard definition. Previous research has attempted to identify colonoscopy indication using ARC [28]. The dataset used, though, was manually selected to contain only pathology reports. The ARC model achieved sensitivity of 0.77 and specificity of 0.88 in classifying pathology reports. The manual selection of documents is only feasible for small projects and does not support ongoing monitoring of colonoscopy indications.

This paper describes the development of rule-based and machine-learning based models for classifying clinical documents with the type of indication for colonoscopy. The performance of these models is also compared with the use of ARC.

Methods

Manual Annotation

In order to create a system that is able to handle a continuous stream of heterogeneous data, the study cohort included all patients who underwent a colonoscopy procedure at the VA facilities in fiscal years 2009 to 2012, as identified using administrative procedure codes. We created a set of metadata filters using an iterative process of manual review to exclude notes with extremely low probability of being a colonoscopy procedure note. These filters were tuned to be inclusive of notes with any indication of relevance and were designed to be scalable for future large-scale implementation of the system. The filters excluded notes authored by staff of unrelated specialties, and included only notes created within a reasonable time frame of the procedure. From this enriched corpus, 2,000 notes were selected for manual annotation, ensuring that all VA facilities that perform colonoscopies were represented in the sample. Each document was independently reviewed by 2 chart abstractors and classified as "Screening", "Non-screening", "Non-colonoscopy", or "Unknown" using an open-source annotation tool eHOST [29]. Reviewers annotated the text span that was most informative in determining the classification. The annotators achieved inter-annotator observed agreement of 0.76 across all classes and 0.79 on Screening. Annotation disagreements were adjudicated by a gastroenterologist. Annotator inconsistency was noted when the same phrase was marked as being indicative of different categories. The conflicting documents were reviewed again by the gastroenterologist, and the

classification discrepancy was resolved. In addition, 1,000 documents were selected randomly from the enriched corpus and manually classified for validation. Table 1 presents the distribution of classifications in the training and validation sets.

Algorithms

The requirements of the project led us to develop a rule-based system as the core of a robust and accurate colonoscopy indications classification system. In order to compare performance of the final system to other approaches, though, we used the same annotated dataset to develop machine learning models.

Table 1 - Class distribution in the training dataset of 2,000 documents and validation dataset of 1,000 documents

Document label	Training Dataset		Validation Dataset	
	Count	Proportion	Count	Proportion
Screening	417	0.209	182	0.182
Non-screening	1018	0.509	388	0.388
Non-colonoscopy	526	0.263	407	0.407
Unknown	39	0.020	23	0.023

Rule-based model (RB)

Each document in the reference standard contained one annotation that reflected the human perception of the indication for the colonoscopy procedure. In addition to the class label, these annotations provided a string of text that was sufficient to make the label determination. We reviewed the distinct text strings in the annotated documents and created an initial rule set for classification. These rules were in the form of regular expressions to mark text containing evidence of colonoscopy indications and a logic module that analyzed the matches and applied heuristics to produce the document label. The first iteration system was applied on the reference set of documents and an error analysis was performed. Analysis of the mislabeled documents produced additional rules. Limiting indication phrase searches to only specific sections was determined to improve classification accuracy. Another rule set was added to filter out documents that were not colonoscopy reports. After implementing the additional rules, another error analysis was performed to confirm the improvement in performance. The last iteration focused on improving section segmentation and optimizing classification rules.

The final system was developed as an NLP pipeline based on the VINCI Leo framework, a set of libraries that enables rapid development and more efficient utilization of the Unstructured Information Management Architecture Asynchronous Scaleout (UIMA AS) [30,31].

The system has the following modules:

- Document type detection determines whether the document is a colonoscopy report or not based on mentions of other procedures or keywords such as “planned procedure”, “informed consent”, or “pre-procedure assessment”.
- Indication section identification was implemented as a two-step process and detects the beginning and the end of sections likely to contain colonoscopy indication if the document is a colonoscopy report. Keywords such as “indication”, “reasons for colonoscopy”, and “preoperative diagnosis” were used to find the beginning, and “consent”, “complications”, or double blank line to find the end of the section.
- Screening indication detection uses keywords and phrases within indication sections that are common to documents classified as “Screening”.
- Non-screening indication detection uses keywords and phrases within indication sections that are common to documents classified as “Non-screening”.

The rule-based classification labeled documents based on the presence of identified keywords and phrases. If a document contained evidence that it was not a colonoscopy report, the document was labeled as “Non-Colonoscopy”. If a document was determined to be a colonoscopy report, contained an indication section, and contained any mention of increased risk, it was labeled with the “Non-screening” class. If the document contained only mentions of average risk, the document was labeled with the “Screening” class. If the document contains an indication section, but the system did not find any evidence that the colonoscopy report was either for screening or non-screening purposes, the document was labeled as “Unknown”.

The system can accept narrative text either from a database or as text files and can output the document names and their corresponding labels to a database or a comma-delimited file. This rule-based approach will be referred to as RB.

Full document SVM model (ML1)

The first supervised machine-learning model was developed using a bag-of-words feature set. Bag-of-words is a commonly used classification approach that minimizes the need to perform expert-driven feature selection by using the presence of the specific words in a document. The feature set in these models are high-dimensional, because each feature stands for a distinct lexical form found in the corpus. The feature sets for this approach used the words from the entire document, which resulted in an average of 286 features per document. A support vector machine (SVM) is a machine-learning algorithm capable of processing large feature sets in multi-class classification tasks. LIBSVM, an open-source Java-based implementation of the SVM algorithm, was used to train a document classifier [20].

In order to use machine learning capabilities of the VINCI Leo architecture, a pipeline was developed to transform documents into their bag-of-words feature representation, to apply LIBSVM on the resulting feature set, to store the SVM model, and to apply that model on a new set of documents. The model created by this system will be referred to as ML1.

Limited context SVM model (ML2)

A second supervised machine-learning algorithm was developed to focus just on the indication section. By narrowing the feature set to only words contained in the indication section, algorithm efficiency and, possibly, accuracy could improve. The indication section identification functionality of the rule-based system was first applied to the document and a second SVM model was created that used a bag-of-words approach only of those words within the indication section. The size of the limited context feature set was reduced to an average of 9 features per document. The model created by this system will be referred to as ML2.

ARC

In order to compare expert-driven and computationally driven feature selection, we decided to use ARC to determine the best feature set for classification. One of the main limitations of ARC is that it is designed to automatically create binary classification models (not multi-class models). Therefore, we had to modify the classification problem to “Screening” versus all other categories. Applying ARC on the reference dataset resulted in a number of models that had various performances

based the selected features sets. Training ARC and determining the best feature set took over 14 hours to complete on the 2,000 training documents. The best-performing feature set was selected for the final model. The systems performances were measured using the standard definitions for document classification tasks as outlined in the figure below.

		Reference standard		
		Screening	Other	Accuracy'='(TP+TN)/'Total
System output	Screening+	True positive (TP)	False positive (FP)	PPV'='TP/(TP+FP)
				NPV=TN/(TN+FN)
	Other	False negative (FN)	True negative (TN)	Sensitivity'TP/(TP+FN)
				Specificity'TN/(FP+TN)

Figure 1 – Formulas used for performance analysis

Results

The models were run on the training dataset, and the training performance is presented in Table 2. ARC failed to process 3 files and they were excluded from the analysis. The four approaches were also applied to the validation set, and the standard performance metrics are presented in Table 3. The binary classification ("Screening" vs. all other labels) is presented in order to compare the models that allow multi-class comparisons with ARC.

Table 2 – Performance of four models on "Screening" versus other labels classification on training data

	RB	ML1	ML2	ARC
True positive	397	411	406	333
True negative	1537	1579	1574	1478
False positive	46	4	9	107
False negative	20	6	11	82
Accuracy	0.967	**0.995**	0.990	0.906
PPV	0.896	**0.990**	0.978	0.757
NPV	0.987	**0.996**	0.993	0.947
Sensitivity	0.952	**0.986**	0.974	0.802
Specificity	0.971	**0.997**	0.994	0.932

Table 3 – Performance of four models on "Screening" versus other labels classification on validation set

	RB	ML1	ML2	ARC
True positive	161	119	152	144
True negative	792	793	786	785
False positive	26	25	32	33
False negative	21	63	30	38
Accuracy	**0.953**	0.912	0.938	0.929
PPV	**0.861**	0.826	0.826	0.814
NPV	**0.974**	0.926	0.963	0.954
Sensitivity	**0.885**	0.654	0.835	0.791
Specificity	**0.968**	0.969	0.961	0.960

We used the standard definitions of classification performance metrics. Because of a narrow and fairly straightforward task definition (and large number of training cases), all four systems performed with a similar level of accuracy, but the rule-based system slightly outperformed others. The specificity of all four systems was not significantly different. However, sensitivity of the full document bag-of-words SVM model of 0.654, compared to the sensitivity of the rule-based system of 0.885, represents a significant difference ($p<0.01$). As the result, the overall accuracy of the rule-based model on test set was also significantly better than the accuracy of ML1 model. In addition to measuring accuracy, we also recorded the processing time required to evaluate the test dataset by each of the models. Processing with ARC took over one hour, and the other three models processed the 1,000 documents in 2 to 3 minutes.

Discussion

The overall system accuracy is only one of many factors to be considered when deciding what computational approach to use for the design of a document classification system. Depending on the purpose of the system, requirements for scalability, flexibility, processing efficiency, and available skill set of the research personnel have to be taken into account. We aimed to create a system capable of processing a large number of files and flag them with one of three labels. This goal set the following system requirements: 1) The system must be able to handle multi-class labeling; 2) The system must be able to process a large number of documents; and 3) The system must be relatively easy to update and maintain.

The demand for manual effort spent annotating reference standard, creating classification heuristics, and engineering a feature set varies by approach, and it must be considered. The narrow problem scope and the restricted nature of the clinical subdomain limited the language variability, making it feasible to develop classification rules manually.

The effect of feature selection on the performance of machine learning system emphasizes need for the thoughtful feature selection. The model that performed the best on the training set was the SVM model created on the full document ML1, achieving almost perfect accuracy. However, the same model performed the worst on the testing set and achieved significantly lower level of accuracy as compared to the training set ($p<0.01$). The main reason for such a large change in performance from training to testing datasets is overfitting. The ML1 model captured the idiosyncrasies of the training set notes and their distributions, but may have encoded features that are not meaningful for the overall problem. Even simple but meaningful feature selection is able to substantially improve performance of a machine learning model as illustrated by the difference in performance of ML1 and ML2 models on the validation set.

While ML2 and ARC both performed feature selection and achieved comparable level of accuracy, the amount of time consumed by ARC to perform classification was substantially longer. The reason for this additional processing time is that ARC runs documents through the default pipeline of cTAKES which results in a large number of features being generated that may not be used for classification.

Limitations

The main goal of the model comparison performed for this study was to compare the performance of the rule-based NLP system to performance of other models that could be created with a minimal investment of time. As a result, we created three machine learning models with different feature sets. Testing results between the rule-based system and the three machine learning systems show performance increase when expert time was invested in rule development. Performance comparison between ML1 and ML2 illustrates performance increase when expert-driven feature selection was performed at the cost of expert time to engineer the feature set. While the ARC model also employed feature selection, ARC's feature set was completely computer generated. ARC performance can be compared to the rule-based NLP system performance to illustrate the performance trade off when using a general-purpose tool as compared to a project specific custom tool.

Conclusions

NLP can be successfully used to detect colonoscopy procedure indications even in case of unknown document type. While machine learning systems are commonly regarded as the holy grail of computational analysis of text, achieving acceptable performance and interpreting the results is challenging. The results of this study indicate that narrow domain and limited scope problems are well suited to be evaluated using heuristic rule-based models rather than machine learning algorithms without thoughtful feature engineering.

Availability of off-the-shelf machine learning tools, such as ARC, makes these methods attractive for research projects. It may be in the combination of off-the-shelf and project specific customization using which the efficiency and accuracy can best be optimized.

Acknowledgement

The primary funding source for this study was QUERI RRP 12-184. Dr. Saini's work is funded by VA HSR&D CDA 09-213. The VA Salt Lake City Health Care System and VA Informatics and Computing Infrastructure (VINCI), VA HSR HIR 08-204, provided resources and facilities. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

[1] Rawl SM, Menon U, Burness A, Breslau ES. Interventions to promote colorectal cancer screening: an integrative review. Nurs Outlook. 2012;60(4):172–181.e13.
[2] Weizman A V, Nguyen GC. Colon cancer screening in 2010: an up-date. Minerva Gastroenterol Dietol. 2010;56(2):181–188.
[3] Sheffield KM, Han Y, Kuo Y-F, Riall TS, Goodwin JS. Potentially inappropriate screening colonoscopy in Medicare patients: variation by physician and geographic region. JAMA Intern Med. 2013;173(7):542–550.
[4] Fisher DA, Maple JT, Ben-Menachem T, et al. Complications of colonoscopy. Gastrointest Endosc. 2011;74(4):745–752.
[5] Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med. 2009;150(12):849–857, W152.
[6] Goodwin JS, Singh A, Reddy N, Riall TS, Kuo Y-F. Overuse of screening colonoscopy in the Medicare population. Arch Intern Med. 2011;171(15):1335–1343.
[7] Ko CW, Dominitz JA, Neradilek M, et al. Determination of Colonoscopy Indication From Administrative Claims Data. Med Care. 2012.
[8] El-Serag HB, Petersen L, Hampel H, Richardson P, Cooper G. The use of screening colonoscopy for patients cared for by the Department of Veterans Affairs. Arch Intern Med. 2006;166(20):2202–2208.
[9] Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. Dig Dis Sci. 2010;55(6):1721–1725.
[10] Lieberman D, Nadel M, Smith RA, et al. Standardized colonoscopy reporting and data system: report of the Quality Assurance Task Group of the National Colorectal Cancer Roundtable. Gastrointest Endosc. 2007;65(6):757-766.
[11] Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA--U.S. Department of Veterans Affairs national-scale HIS. Int J Med Inform. 2003;69(2-3):135–156.
[12] Brown SH, Lincoln M, Hardenbrook S, et al. Derivation and evaluation of a document-naming nomenclature. J Am Med Inform Assoc. 2001;8(4):379–390.
[13] Zeng QT, Redd D, Divita G, Jarad S, Brandt C, Nebeker JR. Characterizing Clinical Text and Sublanguage: a Case Study of the VA Clinical Notes. J Heal Med Informat. 2011;S3.
[14] Baharudin B, Lee LH, Khan K. A Review of Machine Learning Algorithms for Text-Documents Classification. J Adv Inf Technol. 2010;1(1):4–20.
[15] Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv. 2002;34(1):1–47.
[16] Domingue J, Fensel D, Hendler JA, eds. Handbook of Semantic Web Technologies. 1st ed.; 2011.
[17] Mostafa J, Seki K, Ke W. Beyond Information Retrieval: literature mining for biomedical knowledge discovery. In: Biological data mining; 2009:449–480.
[18] Uzuner O, Zhang X, Sibanda T. Machine learning and rule-based approaches to assertion classification. J Am Med Inform Assoc. 2009;16(1):109–115.
[19] Friedman C, Kra P, Rzhetsky A. Two biomedical sublanguages: a description based on the theories of Zellig Harris. J Biomed Inform. 2002;35(4):222–235.
[20] Chang C-C, Lin C-J. {LIBSVM}: A library for support vector machines. ACM Trans Intell Syst Technol. 2011;2(3):1–27.
[21] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explor Newsl. 2009;11(1):10-18.
[22] Zhang Y, Jin R, Zhou Z-H. Understanding bag-of-words model: a statistical framework. Int J Mach Learn Cybern. 2010;1(1-4):43–52.
[23] Forman G. An extensive empirical study of feature selection metrics for text classification. J Mach Learn Res. 2003;3:1289–1305.
[24] Moens M-F. Evaluation of Information Extraction Technologies. In: Information Extraction: Algorithms and Prospects in a Retrieval Context.Vol 21. Springer Netherlands; 2006:179–197.
[25] Barakat N, Bradley AP. Rule extraction from support vector machines: A review. Neurocomputing. 2010;74(1-3):178–190.
[26] D'Avolio LW, Nguyen TM, Farwell WR, et al. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (ARC). J Am Med Inform Assoc. 2010;17(4):375–382.
[27] Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;17(5):507–513.
[28] Hou JK, Chang M, Nguyen T, et al. Automated identification of surveillance colonoscopy in inflammatory bowel disease using natural language processing. Dig Dis Sci. 2013;58(4):936–941.
[29] South BR, Leng J, Anderson K, Shen S, Thibault J, DuVall S. The Extensible Human Oracle Suite of Tools (eHOST) for Pre-Annotation of Clinical Narratives. In: BioCreative: Critical Assessment of Information Extraction in Biology Annual Conference. 2010.
[30] Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat Lang Eng. 2004;10(3-4):327-348.
[31] Cornia R, Patterson OV, Ginter T, Duvall SL. Rapid NLP Development with Leo. AMIA Annu Symp Proc. 2014.

Address for correspondence

olga.patterson@utah.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-619

Automatically Expanding the Synonym Set of SNOMED CT using Wikipedia

Daniel R. Schlegel, Chris Crowner, Peter L. Elkin

Department of Biomedical Informatics, University at Buffalo, SUNY, Buffalo, NY, USA

Abstract

Clinical terminologies and ontologies are often used in natural language processing/understanding tasks as a method for semantically tagging text. One ontology commonly used for this task is SNOMED CT. Natural language is rich and varied: many different combinations of words may be used to express the same idea. It is therefore essential that ontologies and terminologies have a rich set of synonyms. One source of synonyms is Wikipedia. We examine methods for aligning concepts in SNOMED CT with articles in Wikipedia so that newly-found synonyms may be added to SNOMED CT. Our experiments show promising results and provide guidance to researchers who wish to use Wikipedia for similar tasks.

Keywords:

Ontology; Natural Language Processing; Terminology

Introduction

Automated understanding of text within the medical domain relies heavily upon the coverage of clinical terminologies. One such terminology, SNOMED CT, has been used extensively for such tasks [1-10]. An observation which has been noted by researchers examining SNOMED CT's coverage of clinical problem lists is that it could benefit from a more expansive set of synonyms [11].

As of the March 2014 English release, SNOMED CT contains 403,465 concepts, organized into several hierarchies covering medicine and medicine-related domains. Each of the 403,465 concepts contains at least two textual representations – one which includes one of a few dozen semantic types for disambiguation purposes, and one without. To these, there are added 230,863 synonyms which generally consist of alternative names, abbreviations, and shortened forms.

Wikipedia [12] is a community-maintained encyclopedia, covering topics in nearly every imaginable domain. It has a large number of articles related to medicine and science [13], and its scientific articles are of similar quality as Encyclopedia Brittanica [14]. Subdomains of its medicine-related articles, such as mental health, have been shown to have similar accuracy when compared to curated web sources [15]. As of December 2014,[1] Wikipedia contained over 4.6 million articles in English. One source of synonyms in Wikipedia is page redirects.

Wikipedia is being used increasingly often in medicine. In certain subdomains it has shown to be useful as a patient education resource [16]. In 2009, 28% of pharmacists reported using Wikipedia for drug information [17]. It has also been shown to be useful in monitoring infectious disease [18]. Wikipedia is being used in medical research with increasing frequency. Google Scholar finds about 20,300 results for the terms "wikipedia" and "medical" since 2013, and only an extra 1,400 since 2010.

Every article in Wikipedia is tagged with one or more of over 900,000 categories, which form a directed graph ("hierarchy"). Unfortunately, Wikipedia's "hierarchy" of categories "is barely useful for ontological purposes" [19]. For example, *Cath lab* is a page in the category *Cardiac procedures*. A cath lab is certainly a *place* where a cardiac procedure may be performed, but it is not itself a cardiac procedure. Indeed the categorization of pages is often more similar to a collection of related topics, rather than a rigorous ontological classification. For this reason, the ontology alignment techniques which have previously been used with SNOMED CT (such as [20, 21]) are not helpful.

We develop a method for automatically matching SNOMED CT concepts to Wikipedia articles based on lexical matches between synonyms from both, using heuristics, and through alignment of useful portions of the Wikipedia category hierarchy with SNOMED CT semantic types. Previous research has categorized Wikipedia articles as being either health-and-clinically related, or not, using the Wikipedia category hierarchy [22].

Elkin et al. [8] have used a set of 2.5 million synonyms created through a knowledge engineering process in the iNLP system. They used word synonyms such as *cancer* for *neoplasm*, then propagated the synonym through all concepts using the original word, creating new term synonyms. So, the synonym *cancer of tonsil* is added for *neoplasm of tonsil*. One issue with this approach is that many of the synonyms may never appear in actual text. Wikipedia redirects, on the other hand, are created because they are believed likely to occur. We are, of course, not the first to extract synonyms from Wikipedia (see [23] for one of the earliest examples), but we believe we are the first to attempt to enhance SNOMED CT synonymy using heuristic based extraction methods with Wikipedia.

Methods

Wikipedia's redirect pages (henceforth, *redirects*) have no content; they only automatically redirect the user to a specific page. Redirects are designed to get the user to the most appropriate page given their search. Redirects may be: alternative names; plurals; closely related words; pointers from adjectives/adverbs to the noun form; less or more specific forms of names; abbreviations; alternative spellings and common misspellings; alternately punctuated forms; alternative capitalizations; and subtopics within a larger article

[1] We use version 20140614 in this study.

[24]. Page redirects appear to be the most common way to derive synonymy from Wikipedia.

There are other methods for deriving synonymy using Wikipedia. Bolded words in an article's lead section are often synonyms according to the Wikipedia Manual of Style [25]. Analysis of links between pages, and more complex linguistic analysis of the page may also be used. In this paper we focus on redirects, but recognize that the other options may be useful, and leave analysis of them to future work.

SNOMED CT concepts are matched with Wikipedia articles through lexical matches. For example, in Figure 1 you can see that SNOMED CT contains the concept *Entire helicis major muscle (body structure)* with two synonyms. One of those synonyms, *Helicis major*, is also in Wikipedia. Wikipedia then contributes two new synonyms: *Musculus helicis major* and *Large muscle of helix*.

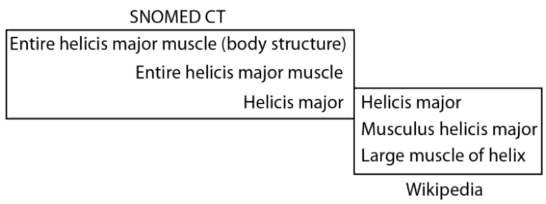


Figure 1 - A match between SNOMED CT and Wikipedia

We first performed an initial evaluation using the naïve matching strategy described above, generating the set of matches between SNOMED CT concepts and Wikipedia articles. We found that simply matching synonyms from SNOMED CT and Wikipedia does not produce very good results (see the Results section). However, analysis of the initial results led to the development of several heuristics, including a mapping between SNOMED CT semantic types and Wikipedia category "hierarchies". We hypothesized the heuristics and the mapping would improve the results.

In the remainder of this section, we discuss some preliminaries about Wikipedia's category hierarchy and what methods we used to make use of it. Then, the problems found with the initial evaluation are presented and solutions proposed. Finally methodology we used in the final evaluation is presented.

Wikipedia Categories

Every page in Wikipedia is a member of one or more categories, defined by the community according to the categorization guidelines [26]. Categories are organized into a directed graph which contains cycles, and includes edges between "hierarchies" from higher to lower level categories. The category graph also allows for multiple inheritance: a subcategory may have more than one supercategory. This structure is illustrated in Figure 2. These characteristics are so prevalent that often the closure of subcategories of an upper level category is all or most of Wikipedia.

To overcome these difficulties with the category graph, we use two independent approaches, with the results later combined. The first approach is naïve – it simply determines if the subcategorical closure of a category is all or most of the categories in Wikipedia. If it is, it recursively navigates up to *d* levels below the initial category, at each step checking again if the subcategorical closure is all or most of the categories in Wikipedia. At the point where the subcategorical closure is not all or most of Wikipedia, or depth *d* is reached, the algorithm cuts subcategory links. For this study, we used *d*=2.

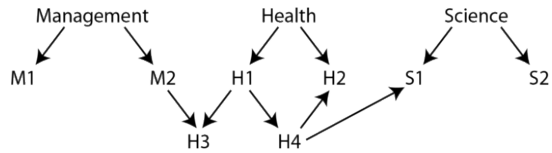


Figure 2 - An illustration of the structure of the Wikipedia category "hierarchy".

The second approach we use is similar to that of [22]: using a breadth-first approach we traverse the graph down from the top-level category (*Articles*) assigning each category a number for its depth. If a category has a subcategory which has a depth less than that of the category itself, that relation is removed.

Alignment

During careful analysis of the initial results, we determined that poor matches and synonyms occurred for five reasons. For each of these we have proposed a solution.

1. Problem: Many incorrect matches between SNOMED CT and Wikipedia (those where none of the resulting synonyms are relevant) occur because a SNOMED CT synonym matches with a Wikipedia redirect for a page outside the appropriate domain.
   Example: The SNOMED CT concept *articular surface of bone* has the synonym *joint*. In Wikipedia, *joint* is polysemous. One use is for drugs, where synonyms such as *dooby* are found.
   Solution: We use the mapping in Table 1 to ensure the domain is maintained. Only the included semantic types are matched on. We require 50% of an article's categories to be in the SNOMED CT semantic type of choice. This value was chosen based on work by [22] categorizing health data in Wikipedia.
2. Problem: Related, but incorrect, redirects often are exact matches of other SNOMED CT terms.
   Example: In Wikipedia, *cutaneous sarcoidosis* redirects to the article *sarcoidosis*. In SNOMED CT, these are two distinct (but related) concepts.
   Solution: Eliminate redirects which match other SNOMED terms from the results.
3. Problem: Acronyms are very polysemous, even within subdomains.
   Example: The acronym *ED* can stand for: eating disorder, effective dose, emergency department, erectile dysfunction, and others.
   Solution: Acronyms are excluded from match criteria, but not results.
4. Problem: If there are more than 10 new synonyms found for SNOMED CT terms outside substances, products, disorders, and observable entities, the new synonyms are often unreliable.
   Example: The SNOMED CT concept *Malus* (the genus for apple trees) when matched against Wikipedia results in 47 new "synonyms", many of which are subtypes like *Malus domesticus*, parts such as *Appleblossom* or even related topics like *apples and teachers*. Out of the 47, only 5 were actual synonyms.
   Solution: These are removed from the results.

Table 1 - Mapping between SNOMED CT semantic types and Wikipedia categories

SNOMED CT Hierarchy	SNOMED CT Semantic Type	Wikipedia Categories
Body Structure	body structure	Anatomy
	cell structure	Cell anatomy
Clinical Finding	finding	Health
	disorder	Health
Geographical location / Environment	geographic location	Geography
	environment	Types of healthcare facilities, Buildings and structures, Human habitats
Event	event	Events
Observable entity	observable entity	Medical signs, Health care
Organism	organism	Organisms
Pharmaceutical / biologic product	product	Drugs, Proteins, Chemical substances, Body fluids
Physical force	physical force	Force, Physical quantities
Physical object	physical object	Physical objects
Procedure	procedure	Medical tests, Health care, Management
	regime/therapy	Medical treatments
Qualifier value	qualifier value	Articles
Record artifact	record artifact	Medical records, Documents, Technical communication
Social context	social concept	Human behavior, Society, Personal life
	ethnic group	Ethnic groups
	racial group	Race (human classification)
Specimen	specimen	Biological specimens, Analytical chemistry
Staging and scales	tumor staging	Cancer staging
	staging scale	Medical scales, Cancer staging
	assessment scale	Medical scales
Substance	substance	Human proteins, Chemical substances

5. Problem: Some subhierarchies of SNOMED CT semantic types contain data not within Wikipedia, and any matches will likely be incorrect.
   Example: The subhierarchy *adjectival modifier* below *qualifier value* contains many adjectives, while adjectives are not well covered by Wikipedia
   Solution: The subhierarchies *adjectival modifier*, and *specific site descriptor* are excluded.

Final Evaluation Methodology

Evaluation was performed again after the solutions in the Alignment section were applied. Two researchers, DRS and PLE, independently created the ground truth used in this study. One-hundred matches and all of their resulting synonyms were randomly sampled and scored. Both annotators classified each new synonym as either being: correct; incorrect, but a subtype; incorrect, but a supertype; incorrect, but related otherwise; or incorrect and unrelated. If a synonym was incorrect but also would never occur in the domain it was excluded from the evaluation results. If a synonym was correct, annotators would classify the synonym as one of: morphological variant – those which a stemmer would find equivalent; spelling variant; capitalization variant; shortened or extended form; eponym; structured coding; or word or term synonym. The number of correct and incorrect synonyms found by DRS and PLE were compared using Cohen's kappa coefficient to measure inter-annotator agreement (κ=.77). Discrepancies between the two annotation results were examined by DRS, and the most correct annotation result was accepted. If a most correct option was not obvious, DRS and PLE discussed the discrepancy until a consensus could be reached.

Results

Initial Evaluation

In this trial we found 43,580 exact lexical matches between SNOMED CT and Wikipedia with 42,958 of those having new synonyms. From these we extracted 446,053 new synonyms. We sampled 100 of the matches (consisting of 988 new synonyms). A single researcher examined these results carefully and found that only 407 (41.2%) of the new synonyms were valid. An additional 360 synonyms (36.4%) were related to the SNOMED CT concept, but were incorrect. These were often related to a higher or lower level concept. The remaining 221 (22.4%) were completely unrelated.

Final Evaluation

After elimination of SNOMED CT concepts from semantic types we do not believe are covered well by Wikipedia (those not in Table 1), there are 272,613 SNOMED CT concepts. Using the heuristics and matching techniques detailed in the methodology section, our system matches 30,781 SNOMED CT concepts. Of those which are matched, 26,580 have new synonyms. Out of the box, SNOMED CT contains 230,863 synonyms. To those, we add an additional 183,100.

Of the 517 synonyms analyzed, we found that 452 (85.6%) of them were correct (true positives), 76 (14.4%) of them were related but incorrect, and 1 (0.2%) was incorrect and unrelated. These percentages are significantly better than the initial evaluation, as visualized in Figure 3.

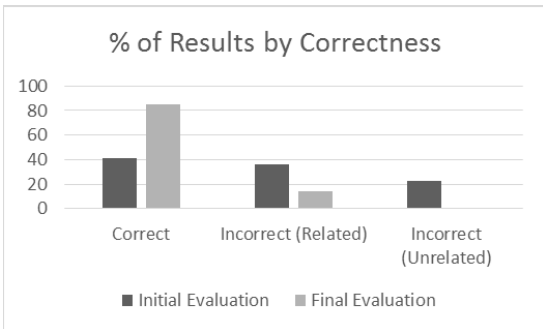


Figure 3 –Percent of correct, incorrect but related, and incorrect and unrelated results in the initial and final evaluations.

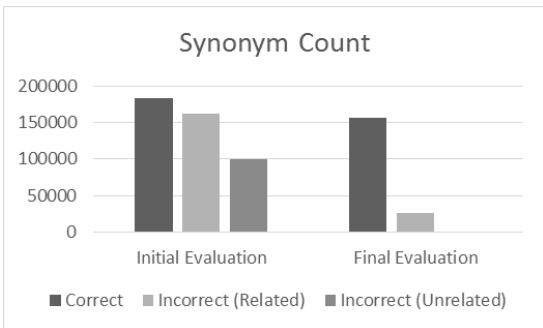


Figure 4 –Correct, incorrect but related, and incorrect but unrelated results in the initial and final evaluations.

Not only did precision improve as a result of our matching technique, few of the correct synonyms found in the initial evaluation were incorrectly eliminated (Figure 4). If we take the sample percentages and apply them to the total number of found synonyms, then compare the initial to final evaluations, we find that the number of correct synonyms decreased from 183,748 to 156,744 (14.7% change), incorrect but related decreased from 162,529 to 26,305 (83.82%), and incorrect and unrelated decreased from 99,775 to 346 (99.65%).

Of the correct results, we found that 61.95% were either word or term synonyms, 4.65% were shortened or extended forms, 0.44% were eponyms, 0.22% were word order variants, 14.60% were various structured codings, and the final 18.14% were capitalization, spelling, or morphological variants. Of the incorrect but related results, we found that 11.84% were subtypes and an additional 24.36% were supertypes. Very few of the synonyms we evaluated would never occur in the domain – only 7 of the total evaluated (1.3%).

Most matches are from the semantic types *body structure* (13.5%), *disorder* (17.9%), *organism* (22.4%), *product* (8.7%), and *substance* (26.4%). Our 100 sampled matches closely followed these percentages. In order to better understand the characteristics of the matching algorithm in these categories, we had an annotator examine random additional examples from these categories until we had around 40 matches from each (see Table 2).

The relative lack of unrelated incorrect results in this table shows that our matching method is correctly matching SNOMED CT terms with Wikipedia articles for the exact concept or a closely related concept.

Table 2 - Statistics for common semantic types

Type	Match Count	Correct Syns.	Incorrect Related	Incorrect Unrelated
body structure	40	126	25	0
disorder	40	174	42	1
organism	40	117	1	1
product	38	628	39	0
substance	40	293	0	0
all others	40	151	85	1

Comparison with Previous Work

We compared the synonyms created using our process with those from Elkin et al.'s iNLP system. We found that only 8,222 of our new synonyms were in their expanded synonym set. We did not use a knowledge engineering approach as they did, which likely would have inflated this number greatly.

Discussion

Using community-sourced data is difficult, because the level of rigor used in its creation is not always of the highest level. The methods we've discussed do not give perfect results, though they are much better than a naïve approach. To help understand what made our system imperfect, we have conducted error analysis on incorrect synonyms.

We have identified three major classes of reasons for incorrect but related synonyms. First, in some cases Wikipedia has redirects from the looked up SNOMED CT term to a related article in which discussion of the concept is only a small part of the article. For example, symptoms of a disease might redirect to a closely associated disease. Consider that *black vomit* redirects to *yellow fever*, since vomit containing blood is a major symptom of yellow fever (and in Spanish, yellow fever is known as *vomito negro* for this reason).

The second reason for incorrect but related synonyms is that a mid-level SNOMED CT concept matches a Wikipedia page which has redirects from more specific or closely associated terms or topics without pages themselves. For example, the page for the genus *Diaptomus* has a redirect from *Diaptomus rostripes* which is a species in that genus with no page itself.

The final reason is that SNOMED CT sometimes has synonyms which are more vague than the concept they stand for. For example, SNOMED CT has a concept for *lower leg*, which is usually defined as the lower extremity from the knee, to the ankle. One of the synonyms is simply *leg*. This is not really a synonym, and that it matches with a more general Wikipedia article should not be surprising.

Incorrect and unrelated terms occurred generally for two reasons. In some cases Wikipedia simply had incorrect redirects. For example *SkyUnion* redirected to *Immunoglobulin G*, while *Sky Union* appears to be a video game company. Some semantic types in SNOMED CT are extremely broad, most notably *qualifier value*. There are many subhierarchies of *qualifier value* which match Wikipedia very well, but some do not. We made an effort to eliminate bad subhierarchies which contained mostly adjectives, but there are others which are problematic. Specifically, it seems that matches from the concepts under *context values* are particularly low quality (*e.g.*, the concept *Done* matches a Wikipedia page about an album named *Done)*. In very broad categories, our matching heuristics have only a small impact.

Some Wikipedia categories such as *Organisms* and *Chemical substances* seem to have articles with very high quality redirects. Chemical substances in particular often have articles for very specific substances instead of redirecting to classes of substances, even if the specific article has little text. Many articles in these categories include codings such as enzyme codes and molecular formulas.

Future research should explore methods to expose deep semantic understanding of articles to both extract more synonyms and ensure correctness. We believe deeper understanding of Wikipedia articles, combined with using parents and children of a SNOMED CT concept being matched, may improve our results. We will also explore methods for attaching provenance and confidence to synonyms. Eventually it is our goal to release frequently updated versions of our synonym set publicly,[2] and use crowdsourcing techniques to continually increase quality.

Conclusion

It is well known that the more synonyms which are available for terminological concepts, the more easily ontologies and terminologies may be used for natural language processing and understanding tasks. We have used Wikipedia redirects as a source to increase the number of synonyms in SNOMED CT by 183,100 with precision of 85.6%. Our techniques for matching SNOMED CT concepts against Wikipedia articles have produced a significant improvement over naïve approaches. Moreover, our experiences with using Wikipedia in this research project may be a valuable resource for other researchers looking to use Wikipedia as an enrichment source.

Acknowledgements

The authors would like to acknowledge Matt Hudson for his helpful comments on previous drafts of this paper.

References

[1] Brown, S.H., et al. eQuality for all - extending automated quality measurement from free text clinical narratives. AMIA Annu Symp Proc. 2008:71-5.
[2] Elkin, P.L., et al., Comparison of NLP Biosurveillance Methods for Identifying Influenza from Encounter Notes. Ann Intern Med, 2012. 156(1 Pt 1): 11-8.
[3] Elkin, P.L., et al. Aequus communis sententia: defining levels of interoperability. Stud Health Technol Inform. 2007;129(Pt 1):725-9
[4] Elkin, P.L., et al. NLP-based identification of pneumonia cases from free-text radiological reports. AMIA Annu Symp Proc. 2008:172-6.
[5] Elkin, P.L., et al., Secondary use of clinical data. Stud Health Technol Inform, 2010. 155: p. 14-29.
[6] Fitzhenry, F., et al., Exploring the fronteir of electronic health record surveillance: the case of postoperative complications. Med Care, 2013. 51(6): p. 509-516.
[7] Garvin, J.H., et al., Automated Quality Measurement in Department of the Veterans Affairs Discharge Instructions for Patients with Congestive Heart Failure. J Healtc Qual, 2012.
[8] Matheny, M.E., et al., Detection of infectious symptoms from VA emergency department and primary care clinical documentation. Int J of Med Inform, 2012. 81(3): p. 143-156.
[9] Matheny, M.E., et al. Detection of blood culture bacterial contamination using natural language processing. in AMIA Annu Symp Proc. 2009: 411-5.
[10] Murff, H.J., et al., Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA, 2011. 306(8): p. 848-855.
[11] Elkin, P.L., et al., Evaluation of the Content Coverage of SNOMED CT: Ability of SNOMED Clinical Terms to Represent Clinical Problem Lists. Mayo Clinic Proc, 2006. 81(6): p. 741-748.
[12] Wikipedia:About - Wikipedia, the free encyclopedia. [cited 2014 Dec. 6]; Available from: http://en.wikipedia.org/wiki/Wikipedia:About.
[13] Friedlin, J. and C.J. McDonald, An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database. J Am Med Inform Assoc, 2010. 17(283e287): p. 283e287.
[14] Giles, J., Internet encyclopaedias go head to head. Nature, 2005. 438(7070): p. 900-901.
[15] Reavley, N.J., et al., Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychological medicine, 2012. 42(08): p. 1753-1762.
[16] Thomas, G.R., et al. An evaluation of Wikipedia as a resource for patient education in nephrology. in Seminars in dialysis. 2013. Wiley Online Library.
[17] Brokowski, L., et al, Evaluation of pharmacist use and perception of Wikipedia as a drug information resource. Ann Pharmacother, 2009. 43(11): p. 1912-1913.
[18] Generous, N., et al., Global disease monitoring and forecasting with wikipedia. PLoS Comput Biol, 2014. 10(11): p. e1003892.
[19] Suchanek, F.M., G. Kasneci, and G. Weikum. Yago: a core of semantic knowledge. in Proceedings of the 16th international conference on World Wide Web. 2007..
[20] Zhang, S., et al. Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference. AMIA Annu Symp Proc. 2005.
[21] Bodenreider, O., Comparing the representation of anatomy in the FMA and SNOMED CT. AMIA Annu Symp Proc. 2006.
[22] Liu, F., et al. Automatically identifying health-and clinical-related content in wikipedia. in MedInfo. 2013.
[23] Milne, D., O. Medelyan, and I.H. Witten. Mining domain-specific thesauri from wikipedia: A case study. in Proceedings of the 2006 IEEE/WIC/ACM international conference on web intelligence. 2006..
[24] Wikipedia:Redirect - Wikipedia, the free encyclopedia. [cited 2014 Dec. 9]; Available from: http://en.wikipedia.org/wiki/Wikipedia:Redirect.
[25] Wikipedia:Manual of Style/Text formatting. [cited 2014 Dec. 9]; Available from: http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Text_formatting.
[26] Wikipedia:Categorization - Wikipedia, the free encyclopedia. [cited 2014 Dec. 9]; Available from: http://en.wikipedia.org/wiki/Wikipedia:Categorization.

Address for correspondence

Daniel R. Schlegel, PhD, Department of Biomedical Informatics, University at Buffalo, 923 Main Street, Buffalo, NY 14203

[2] The synonym set is available for download in the Research section of the UB Biomedical Informatics website: http://www.smbs.buffalo.edu/biomedicalinformatics/.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-624

Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network

Yonghui Wu[a], Min Jiang[a], Jianbo Lei[b], Hua Xu[a]

[a] *School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA*
[b] *Center for Medical Informatics, Peking University, Beijing, China*

Abstract

Rapid growth in electronic health records (EHRs) use has led to an unprecedented expansion of available clinical data in electronic formats. However, much of the important healthcare information is locked in the narrative documents. Therefore Natural Language Processing (NLP) technologies, e.g., Named Entity Recognition that identifies boundaries and types of entities, has been extensively studied to unlock important clinical information in free text. In this study, we investigated a novel deep learning method to recognize clinical entities in Chinese clinical documents using the minimal feature engineering approach. We developed a deep neural network (DNN) to generate word embeddings from a large unlabeled corpus through unsupervised learning and another DNN for the NER task. The experiment results showed that the DNN with word embeddings trained from the large unlabeled corpus outperformed the state-of-the-art CRF's model in the minimal feature engineering setting, achieving the highest F1-score of 0.9280. Further analysis showed that word embeddings derived through unsupervised learning from large unlabeled corpus remarkably improved the DNN with randomized embedding, denoting the usefulness of unsupervised feature learning.

Keywords:

Clinical Natural Language Processing; Named Entity Recognition; Neural Network; Deep Learning; Chinese Clinical Text.

Introduction

The wide use of health information technologies has led to an unprecedented expansion of electronic health record (EHR) data. EHR data have been used not only to support operational tasks in clinical practice (e.g., clinical decision support systems), but also to enable clinical and translational research. However, much of the important patient information is dispersed in narrative clinical documents, which are not directly accessible for computerized applications that rely on structured data. Therefore, clinical natural language processing (NLP) technologies, which can extract important patient information from narrative clinical text, have been introduced to the medical domain and have demonstrated great utility in many applications [1].

Originating from the Sixth Message Understanding Conference (MUC-6) [2], Named Entity Recognition (NER), which aims to identify boundaries and types of entities in text, has been one of the well established and extensively investigated tasks in NLP. In the medical domain, NER for important clinical concepts (e.g., problems, treatments, or lab tests) is also a fundamental step for many clinical NLP systems. Many existing clinical NLP systems use dictionaries and rule-based methods to identify clinical concepts, such as MedLEE [3] - one of the earliest and most comprehensive clinical NLP system developed by Carol Friedman et al. at Columbia University, the MetaMap system [4, 5] - a general biomedical NLP system developed by Aronson et al. at National Library of Medicine, as well as cTAKES [6] - an open source comprehensive clinical NLP system developed based on the Unstructured Information Management Architecture (UIMA) framework and OpenNLP natural language processing toolkit. More recently, a number of challenges on NER involving shared tasks in clinical text have been organized, including the 2009 i2b2 (the Center of Informatics for Integrating Biology and the Bedside) challenge [7] on medication recognition, the 2010 i2b2 challenge [8] on recognizing medical problems, treatments, and tests entities, the 2013 Share/CLEF challenge [9, 10] on disorder mention recognition and normalization, and the 2014 Semantic Evaluation (SemEval) challenge on disorder mention recognition and normalization. For the 2009 i2b2 challenge on medication recognition, although different approaches including rule-based methods, machine learning (ML)-based methods, as well as hybrid methods have been reported, seven out of the top ten ranked systems were rule-based systems that rely on existing biomedical vocabularies to identify medication concepts [7], such as the MedEx system developed at Vanderbilt University [11]. At the 2010 i2b2 challenge for recognizing problems, treatments and lab tests, the organizers provided a large annotated corpus. Therefore, many participating teams including all the top five systems used ML-based methods [12-14]. Many state-of-the-art clinical NER systems implement supervised ML algorithms, such as Conditional Random Fields (CRFs) [15] and Structural Support Vector Machines (SSVMs) [16], with extensive investigation on comprehensive features [17].

In addition to previous NER work on clinical text written in English, there is a growing interest in studying NER of clinical text written in Chinese. With the rapid growth of EHRs implemented in China, there is an urgent need to extract important patient information from Chinese clinical text to accelerate clinical research in China. Conventional ML-based methods have been applied to Chinese clinical NER tasks. Wang et al. [18] conducted a study using CRF, support vector machines (SVM), and maximum entropy (ME) to recognize symptoms and pathogenesis in Chinese medical records. Another study by Wang et al. [19] investigated CRF and different feature sets for recognizing symptom names from clinical notes of traditional Chinese medicine. In 2004, Xu et al. [20] proposed a joint model that integrates segmentation and NER simultaneously to improve the performance of both tasks in Chinese discharge summaries. A more recent work by Lei et al. [21] compared different machine learning algorithms and various types of features for NER in Chinese admission notes and discharge summaries. In summary, current efforts on NER in Chinese clinical text primarily focus on investigating different machine learning algorithms or

optimizing combinations of different types of features via human engineering.

Recently, there is increasing interest in designing deep learning based NLP systems that could automatically learn useful feature representations from large-scale unlabeled corpora through unsupervised learning [22-24]. Deep learning [25, 26] is a research area of machine learning that can learn high-level feature representations by designing deep neural networks. It has achieved state-of-the-art performances in a number of different applications across multiple domains, such as image processing [27], automatic speech recognition [28] and machine translation [29]. Instead of spending a great amount of time on selecting task-specific features, NLP researchers have developed deep neural networks (DNNs) to automatically learn useful features from vast amounts of unlabeled data. Researchers have shown that deep learning approaches can learn useful linguistic features as well as capture semantic meanings through word embedding [30], hence improving the performances of a number of NLP tasks in general English domain. A DNN-based system developed by Dr. Ronan Collobert [23] successfully achieved the state-of-the-art performances on a number of NLP tasks, including POS tagging, Chunking, NER and Semantic Role Labeling, using only one single deep neural network.

In this study, we propose to investigate the use of deep neural network in NER from Chinese clinical text. We developed a deep neural network (DNN) approach for NER in Chinese clinical text and compared it with the traditional CRF-based NER system at the minimal feature setting. To the best of our knowledge, this is the first study to investigate deep neural network in Chinese clinical NER.

Methods

Datasets

Two Chinese clinical corpora were used in this study. The first dataset is an annotated corpus from a previous study by Lei et al. [21], which contains 400 randomly selected admission notes from the EHR database of Peking Union Medical College Hospital in China. For each admission note, four types of clinical entities - problems, lab tests, procedures, and medications were annotated by following the annotation guidelines developed in the study. Details about creation of this dataset can be found in Lei et al. [21]. In summary, a total of 24,433 problems, 2,171 procedures, 11,168 tests and 1,201 medications were annotated in 400 Chinese admission notes. We further divided the 400 admission notes into a training set of two-thirds (266) of the notes and a test set of one-third (134) of the notes. Table 1 shows the distribution of the clinical concepts among the training and test set.

Table 1 – Description of annotated Chinese admission notes

	Training	Test
# Notes	266	134
# Sentences	20,506	10,287
# Characters	277,701	139,885
# Problems	16,253	8,180
# Procedures	1,500	671
# Tests	7,414	3,754
# Medication	840	361

Another dataset, which includes 36,828 unlabelled admission notes (383 Megabytes in total) from the same institute in China, was used for learning word embeddings. Only minimal pre-processing steps (e.g., removing the empty lines) were applied to these notes. As we trained the embedding matrix using individual Chinese characters, word segmentation was not included in the preprocessing.

Experiments and evaluation

In this study, we compared three different NER approaches: 1) the traditional CRF-based NER method, as described in Lei et al. [21]; 2) a DNN-based NER method that uses a randomly initialized word embedding matrix; and 3) another DNN-based NER method that uses the word embedding matrix derived from the unlabelled corpus. All three models were trained using the training set and their performance on the test set are reported. Similar to other deep learning studies, we evaluated the three NER systems at the minimal feature setting, which uses only the word feature. The details of the CRFs model and the DNN models are introduced in the following subsections.

We used the standard micro-average precision, recall and F1-score to evaluate all the three NER systems. All scores were calculated using the Conll 2000 challenge official evaluation script[1]. Wilcoxon signed-ranks test was used to test the statistical significance between two classifiers.

CRF vs. DNN Appraoches

CRF-based NER

The CRFs model decodes the sequence labeling problem by undirected Markov chain and Viterbi algorithm with a training criteria of maximizing the likelihood estimation of conditional probability of the output variable y given the observation x. Here, the observations are the word and its context words in the sentence and the output is its label (e.g, B, I, or O: B - beginning of an entity; I - inside an entity; O - outside of an entity). CRFs was intrinsically designed for sequence labeling problem as it models the relationships between neighboring tokens in sequence. Thus, it has been widely used in various NER tasks and has achieved state-of-the-art performances in both open domains and the biomedical domain. Therefore, CRFs serves as a strong baseline for comparing with the DNN-based NER approaches. More specifically, we used CRF++ package, one of the most popular implementations of CRF (http://crfpp.googlecode.com/svn/trunk/doc/index.html). The parameters for CRFs were optimized using the training set. The details of the CRF-based approach implemented in this study can be found in Lei et al. [21].

DNN-based NER

Researchers have proposed different approaches of designing deep neural networks for NLP systems. In this research, we adopted one of the popular architectures from Dr. Ronan Collobert – the sentence level log-likelihood approach [23], which consists of a convolutional layer, a non-linear layer using the hard version of the hyperbolic tangent (HardTanh), and several linear layers. This structure has been widely used in various NLP tasks and achieved state-of-the-art performances. Figure 1 shows the DNN architecture as well as the propagate function for each layer.

When calculating the classification score for a word, the context words within a specific window size W of the target word are taken as inputs. For the words near the beginning or the end of a sentence, a pseudo padding word will be used to form a fixed length input vector. Each word in the input window can be mapped to an N-dimension vector (N is the embedding dimension) using an embedding matrix. Then, a convolutional layer generates the global features represented as a number of global hidden nodes. Both the local features and the global features are then fed into a standard affine

[1] Available at http://www.cnts.ua.ac.be/conll2000/chunking/conlleval.txt

network trained using back propagation. The lost function is defined using the following sentence level log likelihood:

$$S(X_1^M, T_1^M) = \sum_{t=1}^{M} [H(T_{t-1}, T_t) + DNN(X_t, T_t)] \quad (1)$$

where, $S(X,T)$ is the sentence level log-likelihood score that the sequence of tag T was asigned to the input sequemce X, $H(T_{t-1},T_t)$ is a global transition score from tag T_{t-1} to tag T_t, and the $DNN(X_t, T_t)$ is the score of asigning the tag T_t to an input word X_t assigned by the DNN.

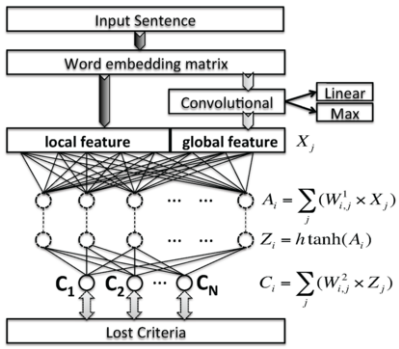


Figure 1 - The sentence approach DNN

Word embedding is a popular method to enrich the traditional bag-of-word representation through mapping the words into real value vectors. Previous research shows that the embedding space is more powerful than the one-hot representation (e.g., bag-of-words) as it conveys more semantic meanings. We adopted the ranking-based embedding method developed by Collobert. The ranking-based embedding treats a sequence of words naturally occurring in the free text as a positive sample. For example, we can form a fixed length positive sample $X=\{w_{L2}, w_{L1}, w_0, w_{R1}, w_{R2}\}$ *for* each word in a sentence given the window size of 5, *where,* w_0 *is the target word,* w_{R1} *and* w_{R2} *are the right-side context words,* w_{L1} *and* w_{L2} *are the left-side context words.* The embedding procedure will generate a negative sample $X^*=\{w_{L2}, w_{L1}, w^*, w_{R1}, w_{R2}\}$ by replacing the central word (w_0) with another word (w^*) that is randomly picked from the vocabulary and try to minimize the ranking criteria with respect to:

$$MAX\{0, 1 - DNN(X) + DNN(X^*)\} \quad (2)$$

The DNN parameters were updated following the standard stochastic gradient descent, as shown in equation 3.

$$\theta = \theta - \lambda \Delta_\theta \quad (3)$$

Where, λ is the learning rate and Δ_θ is the gradient.

We implemented two DNN-based NER approaches for Chinese clinical text. The first model starts with a random initialized word embedding matrix, which will then be updated during the back propagation training procedure. The other DNN model starts with the word embedding matrix derived from the unlabeled notes. The DNN parameters are tuned by splitting one-fifth of the training samples as a validation set using early stopping strategy. Following the work by Collobert [23], we fixed the learning rate at 0.01 and the word embedding dimension at 50. The hidden node number was set to 100, as we tested the numbers from 50 to 500 and noticed that 100 achieved the best performance on the validation set and no further significant improvement was observed when increasing the number of hidden nodes. The window sizes for training the word embedding and sequence labeling were fixed at 11 and 5, respectively. All DNN parameters were updated using the standard stochastic gradient descent through back propagation.

Results

Table 2 shows the performance on the test dataset for the CRFs, the DNN with random initialized word embedding, and the DNN with word embedding derived from the unlabeled corpus. All evaluation scores were based on exact-matching criteria. At the minimal feature setting, the baseline method (CRFs) achieved an F-1 score of 0.9197. The DNN with randomly initialized word embedding achieved an F1-score of 0.9071, which was not better than the baseline performance. The F-1 score of the DNN approach was remarkably improved to 0.9280 by using the word embedding derived from the unlabeled corpus. The best DNN-based NER system outperformed the CRFs by more than 0.8% on F1-score. The Wilcoxon test showed that the best DNN-based NER system outperformed the CRFs with a significant p-value of 1.832e-07. Table 3 shows the performance of the best DNN system for each entity type. The DNN system with word embeddings achieved the highest F1-score of 0.9489 for lab tests and the lowest F1-score of 0.8113 for procedures.

Table 2 – Performances of machine learning methods

	Precision	Recall	F1-score
CRFs	0.9265	0.9130	0.9197
DNN	0.9007	0.9136	0.9071
DNN+Embedding	0.9237	0.9321	**0.9280**

Table 3 – Performances for each entity type (DNN + embedding)

	Precision	Recall	F1-score
Problems	0.9267	0.9356	0.9311
Procedures	0.8119	0.8107	0.8113
Medications	0.8604	0.8366	0.8483
Lab Tests	0.9427	0.9553	0.9489

Discussion

In this study, we examined a novel deep learning method for NER tasks in Chinese clinical text. When only word feature was used, our DNN-based NER system that utilizes word embeddings derived from another unlabeled corpus achieved better performance than the traditional CRF-based NER system, indicating the potential of using DNN for clinical NER in Chinese clinical documents. To the best of our knowledge, this is the first study that investigated deep learning technologies for NER tasks in clinical documents written in Chinese.

We conducted further analysis of the DNN-based NER system and found that the performance improvement was from the semantic information automatically captured by the DNN-based word embeddings. Table 4 shows some examples of semantically related words captured through word embeddings. The nearest neighbors were derived by calculating the cosine similarity using word embeddings. Most of the neighbors in Table 4 are related to the target words, and such semantic relatedness was captured in the embedding space. However, they may not have any relation in the representation space like bag-of-words.

Table 4 – Examples of nearest neighbors captured in word embeddings

一 (one)	左 (left)	肢 (extremity)	喉 (larynx)
三 (three)	右(right)	颌(jaw)	颞(temporal)
二 (two)	双(bilateral)	肺(lung)	局(local)
半 (half)	两(double)	臂(arm)	鼻(nose)
0 (zero)	上(upper)	舌(tongue)	窦(sinus)
两(double)	并(accompany)	壁(wall)	胫(tibia)
数(several)	有(have)	述(mentioned)	睑(eyelid)
有(have)	0.	午(noon)	峡(isthmi)
较(compare)	前(front)	显(show)	脚(foot)
0-0	枕(Occipital)	颈(neck)	涕(nasal mucus)
每(each)	下(lower)	臀(buttocks)	髋(hip)

As NER is an extensively studied task in NLP, many types of features, including the orthographic information, the prefix and suffix, part-of-speech (POS) tags, syntax from parse trees, and existing knowledge bases, have been proved to be useful. On the same dataset, the state-of-the-art CRFs-based NER system that was optimized through manual feature engineering could achieve the best F1-score of 0.9353 [21], which was higher than the DNN-based NER developed in this study. However, the DNN-based NER approach uses word feature only and it requires a minimal effort on feature engineering. The simplicity of the DNN-based approach not only reduces human time on feature engineering, but also makes the implementation of NLP systems easier. For example, obtaining syntax features by parsing sentences could be time consuming and sometimes not feasible in certain real-time applications. However, by using DNN-based approaches, an NER system could still achieve reasonable performance even without running parsers to get the syntax features. Moreover, we believe that the unsupervised feature learning from large unlabeled corpora would significantly reduce NLP researchers' time on task-specific feature engineering, thus enabling efficient development of NER systems for different tasks in various domains.

In the future, we plan to investigate different approaches to further improve the performance of DNN-based NER systems for Chinese clinical text. A straightforward approach is to increase the size of the unlabeled clinical corpus, as previous work has shown that the DNN model benefited from a larger unlabeled corpus. For example, Collobert et al. used the entire English Wikipedia and the Reuters corpus to derive word embeddings. Another appealing research direction is to integrate other linguistic or domain-specific features with the DNN model built on word embeddings to further improve the performance. A previous study by Wang et al. [31] revealed that low dimension continuous space representation works significantly better in DNN than in the traditional systems (e.g., CRFs). However, high dimension discrete feature representation works better in the traditional systems than in the DNNs. Incorporating the high dimension discrete features into the deep neural network remains a challenge. Recent research from Ma et al. [32] showed that the performance of DNN based POS tagging system can be improved by combining a dense embedding feature based neural network with a discrete feature based neural network using a shared output layer. We will further investigate new methods that could combine the two types of feature representations in the future.

Conclusion

In this study, we investigated deep neural network for NER from Chinese clinical text. Our results showed that DNN outperformed CRFs at the minimal feature setting, achieving the highest F1-score of 0.9280. Further analysis showed that the performance improvement was from the semantic information automatically captured by the DNN-based word embeddings, indicating the usefulness of unsupervised feature learning.

Acknowledgments

This study was supported by grant from the NLM R01LM010681-05 and the National Natural Science Foundation of China 81171426.

References

[1] Meystre SM, Savova GK, Kipper-Schuler KC, et al. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform, 2008; pp. 128-44.

[2] Grishman R, and Sundheim B. Message Understanding Conference-6: A Brief History. pp. 466-471.

[3] Friedman C, Alderson PO, Austin JH, et al. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994: 1(2): 161-74.

[4] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp, 2001; pp. 17-21.

[5] Aronson AR, and Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010: 17(3): 229-36.

[6] Savova Gk Fau - Masanz JJ, Masanz Jj Fau - Ogren PV, Ogren Pv Fau - Zheng J, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. 1527-974X (Electronic), 20100907 DCOM- 20101115.

[7] Uzuner O, Solti I, and Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010: 17(5): 514-8.

[8] Uzuner O, South BR, Shen S, et al. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011: 18(5): 552-6.

[9] Suominen H, Salanterä S, Velupillai S, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. Information Access Evaluation. Multilinguality, Multimodality, and Visualization Lecture Notes. In: Forner P, Müller H, Paredes R, et al, eds. Computer Science. Springer Berlin Heidelberg, 2013; pp. 212-31.

[10] Pradhan S, Elhadad N, South BR, et al. Evaluating the state of the art in disorder recognition and normalization of the clinical narrative. J Am Med Inform Assoc 2014: Aug 21.

[11] Xu H, Stenner SP, Doan S, et al. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010: 17(1): 19-24.

[12] de Bruijn B, Cherry C, Kiritchenko S, et al. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. J Am Med Inform Assoc 2011: 18(5): 557-62.

[13] Jiang M, Chen Y, Liu M, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 2011: 18(5): 601-6.

[14] Torii M, Wagholikar K, and Liu H. Using machine learning for concept extraction on clinical documents from multiple data sources. J Am Med Inform Assoc 2011: 18(5): 580-7.

[15] Lafferty J, McCallum A, and Pereira FC. Conditional random fields: Probabilistic models for segmenting and labeling sequence data, 2001.

[16] Tsochantaridis I, Joachims T, Hofmann T, et al. Large margin methods for structured and interdependent output variables; pp. 1453-84.

[17] Tang B, Cao H, Wu Y, et al. Recognizing clinical entities in hospital discharge summaries using Structural Support Vector Machines with word representation features. BMC Med Inform Decis Mak 2013: 13 (1 Suppl): S1.

[18] Wang SK, Li SZ, and Chen TS. Recognition of Chinese Medicine Named Entity Based on Condition Random Field. J Xiamen University (Natural Science) 2009: 48: 349-64.

[19] Wang Y, Liu Y, Yu Z, et al. A preliminary work on symptom name recognition from free-text clinical records of traditional chinese medicine using conditional random fields and reasonable features. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, 2012; pp. 223-30.

[20] Xu Y, Wang Y, Liu T, et al. Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries. J Am Med Inform Assoc 2014: 21: e84-92.

[21] Lei J, Tang B, Lu X, et al. A comprehensive study of named entity recognition in Chinese clinical text. J Am Med Inform Assoc 2014: 21(5): 808-14.

[22] Turian J, Ratinov L, and Bengio Y. Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden, 2010; pp. 384-94.

[23] Collobert R, Weston J, et al. Natural Language Processing (Almost) from Scratch. J. Mach Learn Res 2011: 12: 2493-537.

[24] Socher R, Chen D, Manning CD, et al. Reasoning With Neural Tensor Networks for Knowledge Base Completion. Advances in Neural Information Processing Systems 26, 2013; pp. 926-34.

[25] Jones N. Computer science: The learning machines. Nature 2014: 505(7482): 146-8.

[26] Hinton GE, and Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006: 313(5786): 504-7.

[27] Krizhevsky A, Sutskever I, and Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012; pp. 1097-105.

[28] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE 2012: 29(6): 82-97.

[29] Deselaers T, Hasan S, Bender O, et al. A deep learning approach to machine transliteration. Proceedings of the Fourth Workshop on Statistical Machine Translation, 2009; pp. 233-41.

[30] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[31] Wang M, and Manning CD. Effect of Non-linear Deep Architecture in Sequence Labeling. Proceedings of the Sixth International Joint Conference on Natural Language Processing, 2013; pp. 1285-91.

[32] Ma J, Zhang Y, and Zhu J. Tagging The Web: Building A Robust Web Tagger with Neural Network. Association for Computational Linguistics (ACL), 2014; pp. 144-54.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-629

Identifying Patients with Depression Using Free-text Clinical Documents

Li Zhou[a,b], Amy W. Baughman[b], Victor J. Lei[a], Kenneth H. Lai[b], Amol S. Navathe[c], Frank Chang[a], Margarita Sordo[a,b], Maxim Topaz[b], Feiran Zhong[b], Madhavan Murrali[d], Shamkant Navathe[d], Roberto A. Rocha[a,b]

[a] *Clinical Informatics, Partners eCare, Partners Healthcare Inc. Boston, MA, USA*
[b] *Division of General Internal Medicine and Primary Care, Brigham & Women's Hospital, Harvard Medical School, Boston, MA,USA*
[c] *Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[d] *School of Computer Science, College of Computing, Georgia Institute of Technology, Atlanta, GA, USA*

Abstract

About 1 in 10 adults are reported to exhibit clinical depression and the associated personal, societal, and economic costs are significant. In this study, we applied the MTERMS NLP system and machine learning classification algorithms to identify patients with depression using discharge summaries. Domain experts reviewed both the training and test cases, and classified these cases as depression with a high, intermediate, and low confidence. For depression cases with high confidence, all of the algorithms we tested performed similarly, with MTERMS' knowledge-based decision tree slightly better than the machine learning classifiers, achieving an F-measure of 89.6%. MTERMS also achieved the highest F-measure (70.6%) on intermediate confidence cases. The RIPPER rule learner was the best performing machine learning method, with an F-measure of 70.0%, and a higher precision but lower recall than MTERMS. The proposed NLP-based approach was able to identify a significant portion of the depression cases (about 20%) that were not on the coded diagnosis list.

Keywords:
Depression; Natural Language Processing; Text Classification; Machine Learning'

Introduction

Nearly 1 in 10 adults in the United States have been reported as suffering from clinical depression [1] and similar percentages are reported worldwide [2]. The associated personal, societal, and economic costs are significant [3]. In the U.S., over 15% of depressed people commit suicide, accounting for 30,000 deaths each year. Annual economic consequences are estimated at $83 billion in the U.S. due to higher healthcare utilization and decreased worker productivity [4]. People who have suffered from chronic diseases such as heart disease or diabetes are at greater risk for depression than the overall population. Importantly, those with both depression and other chronic diseases also have worse health outcomes and significantly increased costs compared to their non depressed counterparts [5]. Recently, depression has also been recognized as a major hospital readmission risk factor, and the 30-day readmission rate for mood disorders has been reported at about 15% [6]. A recent study found that 16% of hospitalized adult patients screened positive for mild depressive symptoms and another 24% for moderate or severe depression [7].

Fortunately, most patients with depression can be reliably diagnosed and successfully treated by non-specialists. Only a fraction of patients with major depression require more complex interventions by mental health specialists [8]. Research shows that about 80% of patients with depression will improve after first-line treatment [9]. However, a significant portion of patients do not receive or benefit from depression treatments; only one third of depression patients receive treatment and as many as half of depression cases will go underreported or under-diagnosed [10-13]. Because many complex medical conditions (e.g., heart failure, coronary artery disease, and dementia) are comorbid with depression, it is not surprising that many patients with depression are missed or under-treated as these other issues may take priority during a patient visit. Systematic methods to identify at-risk individuals have the potential to improve patient well-being and quality of life as well as decrease morbidity and suffering.

One strategy to improve this under-treatment is appropriate documentation of a depression diagnosis in electronic health records (EHRs) to better identify depressed patients; and provide evidence-based treatment in a timely fashion. However, emerging evidence shows that using only structured EHR data (such as diagnoses and medications), which are mostly collected for billing purposes, often identifies only a fraction of all documented depression patients. Rather, much of the depression related information is documented in clinical notes [14]. As a result, we need new strategies to more consistently identify depressed patients, particularly those with comorbid medical conditions. Modern information technology methods, such as natural language processing (NLP) can potentially improve the identification and treatment of depression, as well as enable more accurate quality metrics and support new care models targeted at high-risk patients.

Early evidence suggests that NLP methods may be very effective. Few previous studies have focused on the application of NLP technologies to extract psychiatric and mood disorder information from clinician notes [14-16]. A recent study created and validated an NLP application to extract symptomatic remission and treatment resistance information from outpatient psychiatric provider notes for patients with major depressive disorder [15]. Another pilot study has used Twitter data to identify mental health

symptoms for several conditions, including depression [16]. Fischer et al [14] identified depression among diabetes patients using an NLP-based approach by recognizing depression terms and negations from office notes. The authors compared these extracted cases with administrative database codes and found that almost a third more depression cases were identified through NLP only.

The goal of this study is to identify patients with depression using discharge summaries by applying a general NLP system and machine learning classification algorithms. This work is a part of two greater efforts: one study is using NLP to improve patient problem lists by identifying depression diagnoses from free-text notes; and the other utilizes unstructured data from clinical narratives to identify patients at high risk for hospital readmission, augmenting existing methods based on structured administrative data. Depression is of particular interest because it is a risk factor that can enhance the predictive power of existing models of risk-stratification.

Methods

Our methods are summarized in Figure 1. We first manually reviewed a random sample of discharge summaries, identified depression cases, and created a gold standard with the help of domain experts. We then trained and tested our NLP system (known as MTERMS [17]) and classification algorithms. Finally, we assessed our automated approach by comparing system-generated classification against the gold standard. We also compared depression cases extracted from free-text data versus coded diagnoses.

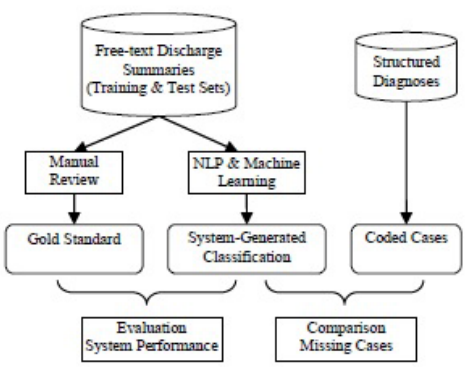


Figure 1-Methods Overview

Data Collection

We used a retrospective cohort of patients from our readmission study. Patients in this cohort had a history of ischemic heart disease. They were hospitalized between 01/01/2011 and 12/31/2013 at different hospitals in Partners HealthCare, a large integrated healthcare network in Boston, Massachusetts, US. We randomly selected 1,200 patients, each with one discharge summary. We then randomly selected 600 discharge summaries for training and used the remaining 600 for testing. At these institutions, discharge summaries include a complete copy of admission note, as well as standard discharge summary information including detailed hospital course, discharge medications, and plan of care.

Identifying Depressed Patients

A pharmacy doctoral student in consultation with an internal medicine physician reviewed both training and testing cases. They classified these cases as depression with high, intermediate and low confidence, based on information in the discharge summary. Twenty randomly selected patient medical records in each classification category were reviewed by one of the clinician authors. This was to validate each category and confirm whether there was enough clinical evidence in the record to support a diagnosis of depression. These findings were then used by the research group to modify our classification algorithms introduced below as needed.

- **High confidence** cases were asserted when depression diagnosis terms were present in notes (e.g., depressive disorder or depression was listed in Past Medical History), indicating a history of depression.
- **Intermediate confidence** cases were identified when combinations of antidepressant treatment (e.g., medications for depression), psychiatry consultation (e.g., referrals to a mental health specialist), or depressive symptomatology were documented. Combinations of above situations can provide clinicians with more information than singular instances when reviewing notes. More specifically, intermediate confidence cases were identified when one of the following scenarios were mentioned in discharge summaries: 1) at least one depressive symptom (e.g., suicidal ideation) and a psychiatric consult was involved (e.g., psychiatry was consulted); 2) at least one symptom of depression was present and an antidepressant medication was prescribed (e.g., citalopram) that did not have an indication in instructions for another problem (e.g., for insomnia); 3) an antidepressant medication was present that did not have an indication in instructions for another problem and a psychiatric consult was involved.
- **Low confidence** cases were asserted when depression diagnosis terms or synonyms were absent and none of the criteria for intermediate confidence case were satisfied. The lack of documentation related to depression cannot exclude a history of depression, but in the scope of discharge notes these were considered as negative or unknown cases.

NLP and Classification Algorithms

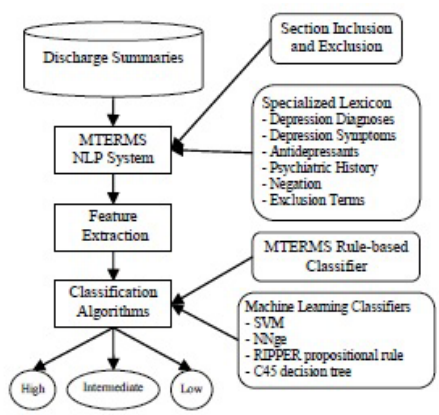


Figure 2 - Overview of NLP and Classification Algorithms

As shown in Figure 2, we first used the MTERMS NLP system [17] to identify relevant sections of the discharge summaries. We included sections that contained medical history, diagnoses, and treatments (e.g., "History of Present Illness", "Social History", "Assessment" and "Medication"). We excluded sections such as "Discharge Instructions", which may mention depression-related terms but do not indicate that the patient actually has depression (e.g., "Call your doctor if you have...suicidal feelings...").

We then created a specialized lexicon containing terms related to depression. For depression diagnoses (e.g., "depression", "depressive"), we used terms that SNOMED-CT (September 2014 Release) classified as preferred terms and synonyms of depression [18]. We identified a list of symptoms (e.g., "insomnia", "suicidal ideation") from criteria listed in Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) [19] and phrases derived from the Patient Health Questionnaire – 9 (PHQ-9) [20], as well as from our previous clinical experience. We used a list of antidepressants (e.g., "amitriptyline", "citalopram") provided by Lexicomp [21], augmented with brand names from the National Library of Medicine's DailyMed website [22]. We also included terms related to a patient's psychiatric history (e.g., "counselor", "psychotherapist"), that we found previously. In addition, we identified terms related to medical electrocardiogram (EKG) findings (e.g., "ST depression") that we found in the training data, since they often contain the word "depression" but are not related to mood disorders. We used MTERMS to extract terms in our specialized lexicon as well as negation terms (e.g., "denies", "no", and "absent").

We then used these terms as features for our classification algorithms, experimenting with both rule-based (within MTERMS) and machine learning classifiers. For machine learning classifiers, we used Weka open-source toolkit [23]. We used 10-fold cross-validation on the training set to select the four best-performing machine learning algorithms to compare on the test set: a support vector machine (SVM) using sequential minimal optimization algorithm [24], a generalized nearest neighbor (NNge) classifier [25], a Repeated Incremental Pruning to Produce Error Reduction (RIPPER) propositional rule learner [26], and a C4.5 decision tree learner [27]. We also manually created a decision tree using Weka's interactive UserClassifier (see Figure 3). At each vertex of the tree, we chose the feature that splits training data most accurately, as long as it was consistent with our clinical knowledge. For example, rules such as "if a symptom is negated, then classify as depression with high confidence" were considered idiosyncrasies of training data, and were therefore not included. We subsequently incorporated our decision tree into MTERMS' classification logic.

System Performance Evaluation

We used standard metrics to evaluate system performance which included precision (p), recall (r), and the F-measure (f). For each category (high and intermediate confidence cases), let TP, TN, FP, and FN be number of true positives, true negatives, false positives, and false negatives, respectively. Then p=TP/(TP+FP), r=TP/(TP+FN), and f=2pr/(p+r). The F-measure is harmonic mean of precision and recall represents their combined quality.

Comparison of Coded vs. Free-text Identification of Depression Patients

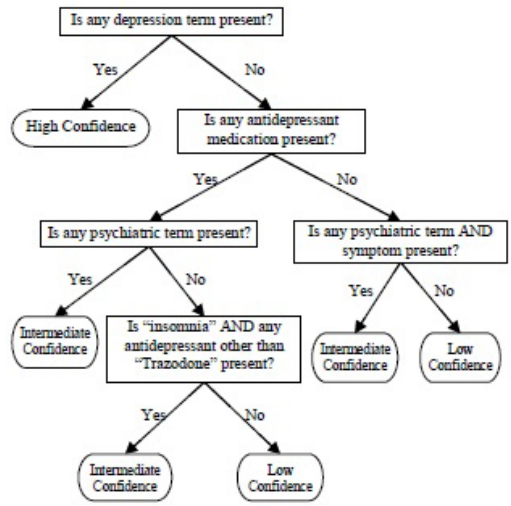


Figure 3- A Knowledge-based Decision Tree

In order to measure how the NLP-based approach in addition to using structured data can enhance the detection of depressed patients, we calculated the percentage of cases that were additionally identified by our NLP system. Coded identification was done with structured diagnosis data which was retrieved from inpatient discharge diagnoses as well as patient problem lists from our outpatient EHR system, when available. We also compared the discrepancies between these data sources.

Results

System Performance

The gold standard established by manual review indicated that out of 600 training cases, 89 depression cases were identified with high confidence and 22 with intermediate confidence; out of 600 test cases, 79 were identified with high confidence and 31 with intermediate confidence. On high confidence cases, all of the algorithms we tested performed similarly; with MTERMS' knowledge-based decision tree slightly better than the machine learning classifiers, achieving an F-measure of 89.6%. MTERMS achieved highest F-measure on intermediate confidence cases, 70.6%. The RIPPER rule learner was the best performing machine learning method, with an F-measure of 70.0%, and a higher precision but lower recall than MTERMS. Full results on our test set are shown in Table 1. The confusion matrix for our knowledge-based decision tree is shown in Table 2.

Comparison of Coded vs. Free-text Data

Using both coded diagnoses and discharge summaries, 140 (23.3%) of the 600 patients in the training set were identified with depression with high confidence, while depression rate in test set was 20.8%, as shown in Table 3.

Even though about 80% of diagnoses were documented in structured data, a significant portion of diagnoses (approximately 20%) were only mentioned in discharge

summaries that were identified by our NLP system. Additionally, for cases identified with intermediate confidence from discharge summaries, more than a quarter (6 of 22 cases in training set and 8 of 31 in test set) were documented in coded diagnosis list, indicating that those intermediate cases that were not on the diagnosis lists may help capture more depressed patients and serve as a good screening candidate for depression.

Table 1- System Performance on Test Set

		High confidence (n=79)			Intermediate confidence (n=31)		
		Precision	Recall	F-measure	Precision	Recall	F-measure
MTERMS decision tree		86.9%	92.4%	89.6%	64.9%	77.4%	70.6%
Machine learning methods	C4.5	87.8%	91.1%	89.4%	64.0%	51.6%	57.1%
	NNge	87.8%	91.1%	89.4%	75.0%	29.0%	41.9%
	RIPPER	85.7%	91.1%	88.3%	72.4%	67.7%	70.0%
	SVM	86.7%	91.1%	88.9%	65.2%	48.4%	55.6%

Table 2-Confusion Matrix for our Knowledge-based Decision Tree

Gold Standard	Knowledge-based Deision Tree		
	High confidence	Intermediate confidence	Low confidence
High confidence	73	2	4
Intermediate confidence	1	24	6
Low confidence	10	11	469

Table 3- Depression Cases Identified using both Coded Diagnoses and NLP

	Coded Diagnoses n (%)	Discharge Summaries & NLP* n (%)	Total Patients with Depression n (%)
Training Set	114 (81.4)	26 (18.6)	140 (23.3)
Test Set	99 (79.2)	26 (20.8)	125 (20.8)

**cases only mentioned in discharge summaries and identified by NLP*

Discussion

This study applied NLP and machine learning classification algorithms to identify clinical depression cases based on discharge summaries. We found a high (over 20%) prevalence of depression among hospitalized patients with a history of ischemic heart disease, consistent with clinical observations that acute myocardial infarction is associated with high risk for on-going depression. Compared to the structured problem list, our automated approach identified about 20% additional depression cases. In Fischer's study [14], 10% of diabetes patients were identified with depression using both administrative data and an NLP method, and an additional 5.3% were identified through NLP alone. While other studies [28] used structured EHR fields (such as billing diagnoses, problem list and medication list) to identify a diagnosis of depression by the primary care physician, our approach can serve as a complementary means to recognize additional cases, particularly in high-risk individuals such as those with ischemic heart disease.

Our study used discharge summaries which are primarily focused on acute medical problems related to the reason for admission. While depression was often clearly documented as a medical problem, we could not always confirm this diagnosis using more comprehensive criteria such as DSM-IV diagnostic criteria or positive results on the PHQ-9 due to limitations in documentation, even when accessing the full medical record. A few previous studies have used psychiatric notes for identifying depression patients. However, most studies do not have access to psychiatric notes, which are often blocked to other providers or are in a separate medical system because of their sensitive content; an algorithm based on this kind of note has limited generalizablity. Our study used a much more pragmatic approach by utilizing a common and readily available type of clinical note; therefore, our methods can be easily replicated within hospitals or primary care practices. By focusing on discharge notes, we also identify patients in a particularly vulnerable time window – after an acute event requiring hospitalization and within the post-discharge period.

Our system achieved an F-measure of 89.6% in identifying high confidence cases and 70.6% for intermediate confidence cases. MTERMS' performance was slightly better than machine learning classifiers. Recall was higher than precision (92.4% vs. 86.9% for high confidence cases and 77.4% vs. 64.9% for intermediate confidence cases), which indicates that such a system will be useful for retrieving relevant instances. Identified cases can then be reviewed by clinicians and researchers. By examining the classification errors made by the system, we found that there were several cases in which our knowledge-based system classified a patient who did not have depression as depressed with either high or intermediate confidence. In some of these false positive cases, a depression-related term was negated, but the negation was outside our NLP algorithm's scope. For example, in the sentence "Negative for fevers / chills / sweats / ... (many other symptoms) ... / depression / ...", the word "depression" was extracted because the negation was too far away. In other cases, depression-related words were used to describe non-mood disorders (e.g., "mild depression of the lateral tibial plateau", "moderate depression of systolic function"). Our lexicon was able to correctly discover some but not all of these phrases. In addition, there were a few cases in which a patient possibly having depression with intermediate confidence was not identified by our system. These false negative cases occurred mainly because the system did not identify symptoms outside our lexicon (e.g., "becomes angry / altered mood"). Our study has a few limitations. First, we only used clinical narrative reports from a single institution's EHR system, thus our results may not be generalizable to other healthcare institutions. We may also have missed terms that physicians use in other institutions. Use of unsupervised algorithms to identify these other terms may be helpful in the future.

Second, we only used discharge summaries. Future work will include applying our system to other types of clinical notes, such as clinic visit notes in outpatient settings. Lastly, patients included in this study had ischemic heart disease and hence results may vary for a general patient population.

Acknowledgments

This study was funded by the Partners Siemens Research Council.

References

[1] Center for Disease Control and Prevention. Current Depression Among Adults - United States, 2006 and 2008. Morbidity and Mortality Weekly Report. Center for Disease Control and Prevention, 2010; 59(38):1229-1235.

[2] Andrade L, Caraveo-anduaga JJ, Berglund P, Bijl RV, Graaf RD, Vollebergh W, Dragomirecka E, Kohn R, Keller M, Kessler RC, Kawakami N, Kiliç C, Offord D, Ustun TB, Wittchen HU. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. International Journal of Methods in Psychiatric Research. 2003;12(1): 3-21.

[3] National Alliance on Mental Illness. The Impact and Cost of Mental Illness: The Case of Depression. National Alliance on Mental Illness (NAMI). Retrieved from: http://www.nami.org/Template.cfm?Section=Policymakers_Toolkit&Template=/ContentManagement/ContentDisplay.cfm&ContentID=19043.

[4] Donohue JM, Pincus HA. Reducing the societal burden of depression. Pharmacoeconomics. 2007;25(1):7-24.

[5] Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. The Lancet. 2007;370(9590):851-858.

[6] Elixhauser A, Steiner C. Readmissions to U.S. Hospitals by Diagnosis, 2010: Statistical Brief #153. Healtcare Cost and Utilization Project. Retrieved from: http://wwwhcup-usahrqgov/reports/statbriefs/sb153pdf.

[7] Cancino RS, Culpepper L, Sadikova E, Martin J, Jack BW, Mitchell SE. Dose-response relationship between depressive symptoms and hospital readmission. Journal of Hospital Medicine. 2014;9(6):358-364.

[8] Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Archives of Internal Medicine. 2006;166(21):2314-2321.

[9] Campbell KP, Lanza A, Dixon R, Chattopadhyay S, Molinari N, Finch RA, editors. A purchaser's guide to clinical preventive services: moving science into coverage. Washington, DC: National Business Group on Health; 2006.

[10]Harman JS, Edlund MJ, Fortney JC. Disparities in the adequacy of depression treatment in the United States. Psychiatric Services. 2004;55(12):1379-1385.

[11]Simpson SM, Krishnan LL, Kunik ME, Ruiz P. Racial disparities in diagnosis and treatment of depression: a literature review. Psychiatric Quarterly. 2007;78(1):3-14.

[12]Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. The Lancet. 2009;374(9690):609-619.

[13]Edlund MJ, Unutzer J, Wells KB. Clinician screening and treatment of alcohol, drug, and mental problems in primary care: results from healthcare for communities. Medical Care. 2004;42(12):1158-1166.

[14]Fischer LR, Rush WA, Kluznik JC, O'Connor PJ, Hanson AM. Abstract C-C1-06: identifying depression among diabetes patients using natural language processing of office notes. Clinical Medicine & Research. 2008;6(3-4):125-126.

[15]Perlis R, Iosifescu D, Castro V, Murphy S, Gainer V, Minnier J, Fava M, Weilburg JB, Churchill SE, Kohane IS, Smoller JW. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychological Medicine. 2012;42(01):41-50.

[16]Harman GCMDC. Quantifying mental health signals in Twitter. ACL 2014, 51.

[17]Zhou L, Plasek JM, Mahoney LM, Karipineni N, Chang F, Yan X, Rocha RA. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. AMIA Annual Symposium proceedings. 2011; 1639-1648.

[18]SNOMED CT. Retrieved from: http://www.ihtsdo.org/.

[19]Diagnostic and statistical manual of mental disorders, 4th ed. American Psychiatric Association, 2000.

[20]Kroenke K, Spitzer RL, Williams JB. The Phq‐9. Journal of General Internal Medicine. 2001;16(9):606-613.

[21]Lexicomp. Retrieved from: http://www.lexi.com/.

[22]DailyMed. Retrieved from: http://dailymed.nlm.nih.gov/dailymed/index.cfm.

[23]Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explorations. 2009;11(1):10-18.

[24]Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing.2004;14(3):199-222.

[25]Martin B. Instance-based learning: nearest neighbour with generalisation. University of Waikato, 1995.

[26]Cohen, WW. Fast effective rule induction. In Machine Learning: Proceedings of the Twelfth International Conference, Lake Tahoe, California, 1995.

[27]Quinlan, JR. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[28]Trinh N-HT, Youn SJ, Sousa J, Regan S, Bedoya CA, Chang TE, Fava M, Yeung A. Using electronic medical records to determine the diagnosis of clinical depression. International Journal of Medical Informatics. 2011;80(7):533-540.

Address for correspondence

Li Zhou, MD, PhD

Clinical Informatics, Partners eCare, Partners HealthCare

93 Worcester Street, Wellesley, MA 02481

lzhou2@partners.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-634

An Approach for Automatic Classification of Radiology Reports in Spanish

Viviana Cotik[a], Darío Filippo[b], José Castaño[a]

[a] *Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina*
[b] *Hospital De Pediatría, Prof. Dr. Juan Pedro Garrahan, Argentina*

Abstract

Automatic detection of relevant terms in medical reports is useful for educational purposes and for clinical research. Natural language processing (NLP) techniques can be applied in order to identify them.

In this work we present an approach to classify radiology reports written in Spanish into two sets: the ones that indicate pathological findings and the ones that do not. In addition, the entities corresponding to pathological findings are identified in the reports.

We use RadLex, a lexicon of English radiology terms, and NLP techniques to identify the occurrence of pathological findings. Reports are classified using a simple algorithm based on the presence of pathological findings, negation and hedge terms.

The implemented algorithms were tested with a test set of 248 reports annotated by an expert, obtaining a best result of 0.72 F1 measure. The output of the classification task can be used to look for specific occurrences of pathological findings.

Keywords:

Natural language processing; Radiology reports; Pathological findings, Negation detection; Text classification.

Introduction

Automatic detection of relevant terms in medical reports is useful for educational purposes, for clinical research and for comparison of findings between institutions.

According to [1], approximately half of the medical conditions described in the medical domain are negated. There also exist hedges (uncertain facts). Being able to differentiate which conditions are present and which are absent in a medical report is a current topic in the area of natural language processing (NLP) [2,3].

We describe here an approach to identify reports containing pathological findings. We work on a set of medical reports of imaging studies (usually called *radiology reports*) in Spanish. Identifying which reports contain pathological findings will allow the indexing of relevant documents only and discard those which are not relevant (do not contain pathological findings).

In order to test the results of our classification algorithm we use a Test Set annotated by a radiology physician, one of the authors of this paper. The Test Set consists of 248 ultrasonography reports that are annotated indicating medical findings and their anatomical location. We obtain an F1 of 0.72, a recall of 0.83 and a precision of 0.63.

We use NLP tools and techniques such as lemmatization, frequency of bigrams and trigrams, part-of-speech tagging (POS tagging), and hedge and negation tagging, in order to process our data. We also used a radiology ontology. Then we tested some simple algorithms to determine whether there is a factual pathological finding in a report.

There exist different ontologies, terminologies and coding systems in the medical domain such as SNOMED CT[1], MeSH[2], ICD-10[3], LOINC[4], UMLS[5] and RadLex[6]. The latter has specifically been developed to satisfy standardized indexing and retrieval of radiology information. It satisfies the needs in this domain by adopting features of existing terminology systems as well as producing new terms to fill critical gaps. It unifies and supplements other lexicons while it also has mappings to them. However, there is no radiology ontology or machine readable dictionary data that can be used to identify terms that denote pathological findings in Spanish. Using a machine translation from an English ontology presents a number of difficulties:

- Some terms frequently used in Spanish with synonyms are less frequently used in English. For example *arteria mamaria interna* for *internal mammary artery* is commonly used in Spanish, while in English it is referred to as *internal thoracic artery*.
- Sometimes adjectives are preferred to nouns in Spanish. For example, *folículo ovárico* for *ovarian follicle* is commonly used, while in English *follicle of ovary* is the preferred term.
- Terms of interest can be composed of more than one word, which often leads to problems in the order of the translated words.

We use RadLex as the main source of information to detect pathological findings.

Given the amount of annotated text is small, it is not possible to use machine learning (ML) techniques to improve the classification algorithm.

Current results enable physicians to quickly detect diagnoses in the reports and, in the future, images related to them. These

[1] Systematized Nomenclature of Medicine Clinical Terms - SNOMED CT, http://www.ihtsdo.org/snomed-ct/
[2] Medical Subject Headings http://www.ncbi.nlm.nih.gov/mesh/
[3] ICD10: International Statistical Classification of Diseases and Related Health Problems 10th Revision, http://apps.who.int/classifications/icd10/browse/2010/en
[4] LOINC. Logical Observation Identifiers Names and Codes. Database and universal standard for identifying medical laboratory observations, http://loinc.org/
[5] Unified Medical Language System, http://www.nlm.nih.gov/research/umls/
[6] Radiology Lexicon: http://rsna.org/RadLex.aspx

results are planned to be used by physicians in a public hospital in Argentina.

Related work

There are several works addressing related problems. There are existing systems process texts in English and there is some work performed for German.

Khresmoi project[7] uses information extraction from unstructured biomedical texts in a cross-lingual environment. They used RadLex.

MoSearch [4], RADTF [5] and Render [6] allow searching for terms in radiology reports taking into account negation and modality information and using NLP techniques. In the last two, results are linked with images from a picture archiving and communication system (PACS). In RADTF, if the user searches for a RadLex term, it returns its RadLex id. They are mainly used for education and research.

RadMiner [7] retrieves images in radiology reports based on NLP techniques. Bretschneider et al. [8] use a grammar-based sentence classifier to distinguish 'pathological' and 'non-pathological' classes. They report 0.74 recall and 0.54 precision measures. Both are implemented for German and use a German available version of RadLex as a linguistic resource. RadMiner adds new terms taken from the annotation performed by a specialist.

MetaMap [9] recognizes UMLS concepts in medical texts written in English.

Bioportal[8], a repository of biomedical ontologies, provides a tool that tags text based on an ontology selected by the user. There are no Spanish ontologies available. A UMLS semantic type can be selected.

LEXIMER [10] uses information theory to classify English radiology reports on the basis of the presence or absence of positive findings. They report precision of 0.98 and recall of 0.99.

Negex[1] is a simple algorithm to identify negations in medical texts written in English. It has been implemented in several languages [3,11,12]. Diverse techniques, such as pattern matching, machine learning and a combination of techniques have been applied to negation identification [2,13,14]. Some challenges have been performed: 2010 i2B2/VA Challenge for clinical text[9], ConLL 2010 for biomedical texts[10], and BioNLP 2009 for biological texts[11].

As far as we are aware of, there are no available systems that identify RadLex terms in Spanish radiology reports.

The rest of the paper is organized as follows. The Methods section presents the approach used in this work as well as data, specific techniques and tools used. The Results section explains metrics and shows current results. Final sections are Discussion and Conclusions.

Methods

The proposed solution is composed of several interconnected but independent modules (see Figure 1).

The *syntactic analysis* module does segmentation, lemmatization, normalization and POS tagging as well as parsing. The *entity recognition* module does dictionary lookup, non-exact recognition, and the *hedge* module identifies hedges and negations. Finally, the classification module classifies texts based on the results of previous modules.

The output of these modules is used with the available data to identify pathological findings that might be of interest for physicians.

In the rest of this section we explain in detail the components of each module and the data used. RadLex data has to be obtained, filtered and translated to Spanish. We explain how we filtered the data, and the translation methods used. We also present the annotation process performed by a specialist, and finally, the methods and techniques used to perform each test.

Figure 1–Modules of the proposed solution

Data

We have about 130,000 medical reports from three different studies: ultrasonography (US), computed tomography (CT) and magnetic resonance imaging (MRI). Table 1 shows the number of available reports of each type.

Table 1– Number of reports available of each type of study

Type of Study	Number of Reports
MRI	14635
CT	29327
US	85621

Reports are in non-structured format (the first part is semi-structured). They are brief (approximately 5 lines each) and they state what was found in the study performed on the patient. An example of an annotated ultrasonography report can be seen in the Annotation section.

RadLex, has different versions[12]. We decided to use version 3.6 since it has improvements over the previous ones and it has been used in other works, such as Bioportal, which allows us to compare results. Furthermore, it is being translated by physicians. Version 3.6 has more than 30,000 terms, that are classified[13], among others, by *imaging modality*, *procedure*, *object*, *imaging observation*, *non-anatomical substance*, *anatomical entity* and *clinical finding*. We selected the terms corresponding to *clinical findings* (what we call pathological findings) and *anatomical entity* type (see the section Technical Details).

Translation

As far as we know, there is no complete RadLex translation to Spanish (the translation mentioned in RadLex reference[14] is partial and not every term is precise). In order to be able to use RadLex with Spanish text we had to obtain a translated version.

All RadLex terms were translated to Spanish with Google Translate[15]. We also used 1) a mapping of RadLex and UMLS terms and through UMLS we obtained the corresponding translation of RadLex terms and 2) a mapping

[7] Khresmoi project: www.khresmoi.eu
[8] Bioportal annotator: http://bioportal.bioontology.org/annotator
[9] https://www.i2b2.org/NLP/Relations/
[10] http://www.clips.ua.ac.be/conll2010/
[11] http://www.nactem.ac.uk/tsujii/GENIA/SharedTask/

[12] RadLex Release Notes: https://docs.google.com/document/d/10zRIBkXyj1eLt3LS_A7w3gSGRedXXoYvucH6H4trCZY/edit?hl=en
[13] Radiology Lexicon: http://rsna.org/RadLex.aspx
[14] Who is using RadLex? https://www.rsna.org/Who_Uses_RadLex.aspx
[15] Google Translate, https://translate.google.com/

of English-Spanish Wikipedia[16] terms. Table 2 shows the number of terms translated using different types of translation sources.

Table 2– Number of English-Spanish RadLex translated terms. The second column refers to the number of RadLex terms translated, and the third column to the number of RadLex terms translated of pathological and anatomical type.

Source of translation	Number of RadLex terms	Anat. and pathological terms
Google Translate	30,000	10,357
UMLS	1304	857
Wikipedia	1620	896

In the Entity Recognition section we explain how we used the translations obtained from different sources.

Preprocessing and syntactic analysis

In the syntactic analysis module all the words of radiology reports were normalized. Freeling[17] was used to perform tokenization and lemmatization. We also used Python[18] to process all radiology reports to obtain frequency of words (unigrams), bigrams, and trigrams.

Entity recognition

In order to detect anatomical and pathological entities, we identified in the reports words that are part of some RadLex term. For example: *vessel* does not appear as a RadLex term but is part of more than 100 RadLex terms (as in *blood vessel*), so the term is included as a *term of interest* to be identified in reports.

Each word appearing in a RadLex term was indexed using an inverted index. Each word in the inverted index points to a set of RadLex terms in which the word occurs. Each RadLex term in this set contains their entity class information, i.e. whether they are anatomical or pathological. Using this information, the entity class is assigned to the indexed word. Then, for each report a decision is automatically made to decide whether the word is a single or multi-word term, and the resulting term is tagged with its entity class. In this step stop words are not considered.

The output of this module is the radiological report with the anatomical and pathological terms automatically annotated according to RadLex terms. A set of common pathological terms compiled by the radiologist is also used to identify *interesting* terms in reports.

All the pathological terms identified in the set of 129,583 reports were stored and the most frequent ones were analyzed. Some of them did not appear to be pathological, so we analyzed bigrams and trigrams containing them and the inverted index in order to check if they were incorrectly tagged as pathological findings. We compiled another dictionary with those terms that we considered to be non-pathological, and we used this dictionary to filter out these terms in the tagging process.

Negation and hedge detection

To detect negated terms and hedge signals we compiled a set of negations and a set of hedges (based on a translation to Spanish of RADTF negations and hedges). These two sets of words were used in a simple dictionary lookup to tag these words in the reports. This is very similar to the approach used by NegEx[1]. If one term is contained in another we get the largest of the two terms, for example if *no* and *no se encontró* are in the negation dictionary and *no se encontró* is in the report we will tag this phrase, rather than the phrase *no*.

Classification

Reports are tagged with pathological entities, negations and hedges. Only those reports that contain positive findings are considered relevant. We defined three simple algorithms in order to determine it.

- Algorithm 1. If there is some pathological finding in the report we identify the report as pathological. It is not taken into into account whether or not there are negations in the text.
- Algorithm 2. If there is a pathological finding identified in the report and there is a negation or hedge somewhere in the report (might be in another sentence), the report is identified as non-pathological.
- Algorithm 3. A report is identified as pathological only if it has at least one sentence with a term indicating a pathological finding and no negation or hedge (in the same sentence).

Annotation

In order to test the results of our classification algorithms, we needed some annotations performed by an expert. We elaborated annotation guidelines stating the criteria to be used for the annotations. A number of annotation-revision iterations were performed until the annotations were as expected. Three experts annotated two sets of 17 and 12 reports. The F-measure of the annotations agreement was 0.7. Once the final version of the annotation guidelines was defined, a radiologist annotated 248 ultrasonography reports with the Callisto[19] annotation tool. These 248 reports were used as a Test Set to evaluate the strategies we implemented. Each report was automatically searched for the presence of *pathological findings* annotations, and based on this it was classified as pathological (if there was at least a pathological finding annotated in the report) or non-pathological (if there was no pathological finding annotated in the report).

An example of an annotation in Spanish and it's translation to English can be seen below:

33289|16a4m|20070807|A27611 HIGADO:<RADLEX> lobulo caudado aumentado de tamano</RADLEX>, resto de higado de ecoestructura conservada. VIA BILIAR intra y extrahepatica: no dilatada. VESICULA BILIAR: alitiasica. Paredes y contenido normal. PANCREAS: tamano y ecoestructura normal. <RADLEX>BAZO: minimamente aumentado de tamano</RADLEX>. Diametro longitudinal:13.5 (cm) RETROPERITONEO VASCULAR: sin alteraciones. No se detectaron adenomegalias. No se observo liquido libre en cavidad. Ambos rinones de caracteristicas normales.

33289 |16y 4m |20070807|A27611 LIVER: <RADLEX> caudate lobe with increased size </ RADLEX>, the other lobes of the liver appear normal. Intra and extrahepatic BILIARY TREE: not dilated. GALLBLADDER: no gallstones were seen. Wall and content appear normal. PANCREAS: normal size and echotexture. <RADLEX> SPLEEN: minimally increased in size </ RADLEX>.

[16] Wikipedia: https://www.wikipedia.org/
[17] Freeling. An open source suite of language analyzers: http://nlp.lsi.upc.edu/freeling/
[18] https://www.python.org/about/
[19] Callisto annotating tool http://annotation.exmaralda.org/index.php/Callisto

Longitudinal diameter: 13.5 (cm) VASCULAR RETROPERITONEAL COMPARTMENT: unremarkable. No lymphadenopathy was detected. No free fluid in the peritoneal cavity was observed. Both kidneys unremarkable.

Technical details

RadLex was downloaded in Protégé format. The selection of anatomical and pathological RadLex has been performed with the help of the tutorial performed by MantasCode[20].

Freeling was used for the syntactic analysis module and Python for implementing the remaining modules.

Results

Table 3 shows the results of evaluation of the three algorithms used to identify reports containing pathological findings against the Test Set. As a reference we describe formulas of calculated metrics. They are accuracy (acc): (TP+TN)/(TP+FN+FP+TN), precision (prec): (TP/(TP+FP)), recall: (TP/(TP+FN)), F1: 2*(prec * recall)/ (prec + recall).

Table 3 – Results of comparison of three algorithms with the Test Set. Algorithm 1: negations are not taken into account. Algorithm 2: negations are taken into account on a report basis. Algorithm 3: negations are taken into account on a sentence basis. References: acc.: accuracy, prec: precision, alg.: algorithm TP: true positives, FN: false negatives, FP: false positives, TN: true negatives.

Measure	alg. 1	alg. 2	alg. 3
acc.	0.60	0.57	**0.67**
prec.	0.56	**0.74**	0.63
recall	**0.96**	0.25	0.83
F1	0.71	0.38	**0.72**
TP	122	32	106
FN	5	95	21
FP	95	11	62
TN	26	110	59

Discussion

Algorithm 3 is the one with best results, since it has the best F1 (a measure that balances precision (those identified as positive and how many are really positive) and recall (the proportion of the positive findings that were retrieved) and the best accuracy (rate of correctly classified documents)). Algorithm 1 has naturally a greater amount of TP, but also of FP, that are decreased with Algorithm 3. This is consistent with the algorithm used because all the findings were tagged independently of the occurrence of negations or hedges.

These results show that there is room for improvement, in particular regarding precision results, and they are promising considering that we are working with very noisy data given that terms used to identify pathological findings were obtained through automatic machine translation. We can assume that as a first step to identify reports with pathological findings, the results are good.

LEXIMER has better results for English and our work has better results than that of Bretschneider et al. for German, but in both cases the results are incomparable, since they have been obtained with different data and for different languages.

[20] JAVA: How to programmatically manipulate a Protégé-Frames lexicon/ontology/dictionary using Protege API and Java. http://mantascode.com/java-how-to-programmatically-manipulate-a-protege-frames-lexicon-ontology-dictionary-using-protege-api-and-eclipse/

Conclusion

Although there are tools that generate structured radiology reports aiming at easier information retrieval, unstructured text is still preferred by most radiologists. It allows a better formulation of their ideas, and writing the report as a continuum, instead of doing it with check boxes and templates [15]. Given that situation, NLP is incorporated as a promising resource for information extraction in this context.

Identifying the frequency of findings and diagnoses found in the different imaging modalities, through the use of information extraction from unstructured radiology reports, should improve aspects of diagnosis and patient care within an institution.

The possibility of linking these findings with corresponding images through PACS makes the radiologist's task easier when he has to evaluate studies and prepare reports. It allows comparison with previous studies that may have similar findings. This set of images and text provide an excellent support for decision making.

We are not aware of existing solutions for Spanish reports. Once the work is finished it could be used in the reports of other Spanish speaking hospitals.

In terms of NLP, the challenges are the application of existing techniques to Spanish, the non-availability of RadLex in Spanish, and the scarcity of resources (annotations) that do not allow us to use ML techniques to improve the classification algorithm.

Future Work

Currently we are working on a number of subjects: 1) improvement of translations (performed by radiologists). This might provide a resource for achieving better entity recognition in the future, 2) enlargement of the manually annotated Test Set. This will allow us to use ML techniques to improve our classification algorithm. The use of boosting is being considered, 3) detection of scope of negation to improve classification (i.e. knowing what is actually being negated). Dependency parsers and ML techniques can be used to identify the scope of negation and hedges. We are also working on the implementations of Negex to Spanish, and 4) evaluation and improvement of detection of findings. We will compare the results of our algorithm with the use of additional resources, such as SNOMED CT and ICD-10 instead of RadLex.

As future work it would also be important to do automatic anonymization of radiology reports. Information about the physician who performed the study and medical record number of the patient should be removed. It is an important task because we are working with sensitive information. Nowadays we are not working with image information from PACS, but we have the keys to relate the reports to their corresponding images. A separate project is being carried out by other people in order to relate the information extracted from reports with the associated images.

References

[1] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics, 2001: 34(5):301-10.

[2] Wu AS, Do BH, Kim J, Rubin DL. Evaluation of Negation and Uncertainty Detection and its Impact on Precision and

Recall in Search. Journal of Digital Imaging 2011, 24(2): 234–242.
[3] Chapman WW, Hilert D, Velupillai S, Kvist M, Skeppstedt M, Chapman BE, Conway M, Tharp M, Mowery DL, Deleger L. Extending the NegEx Lexicon for Multiple Languages. Studies in Health Technology and Informatics, 2013: 192: 677-681.
[4] Ramaswamy MR, Patterson DS, Yin L, Goodacre BW. MoSearch: a radiologist-friendly tool for finding-based diagnostic report and image retrieval. Radiographics, 1996. 16(4): 923-33.
[5] Do BH, Wu A, Biswal S, Kamaya A, Rubin, DL. Informatics in radiology: RADTF: a semantic search-enabled, natural language processor-generated radiology teaching file. Radiographics, 2010: 30(7), 2039-48.
[6] Dang PA, Kalra MK, Schultz TJ, Graham SA, Dreyer KJ. Informatics in radiology: Render: an online searchable radiology study repository. Radiographics, 2009: 29(5):1233-46.
[7] Gerstmair A, Daumke P, Simon K, Langer M, Kotter E. Intelligent image retrieval based on radiology reports. European Radiology, 2012: 22(12): 2750-8.
[8] Bretschneider C, Zillner S, Hammon M. Identifying Pathological Findings in German Radiology Reports Using a Syntacto-semantic Parsing Approach. Proc of Workshop on Biomedical Natural Language Processing, 2013: 27-35.
[9] Aronson A: A Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. Proc AMIA Symp, 2001:17-21.
[10]Dreyer, KJ, Kalra MK, Maher MM, Hurier AM, Asfaw BA, Schultz T, Halpern EF, Thrall JH. Abbreviations : Application of Recently Developed Computer Algorithm for Automatic Classification of Unstructured Radiology Reports. Radiology, 2005, 234: 323–329.
[11] Skeppstedt M. Negation detection in Swedish clinical text: An adaption of NegEx to Swedish. Journal of biomedical semantics, 2011. 2 (S3):S3 doi:10.1186/2041-1480-2-S3-S3.
[12] Costumero R, Lopez F, Gonzalo-Mart C. An Approach to Detect Negation on Medical Documents in Spanish. Brain Informatics and Health. Lecture Notes in Computer Science, 2014. Volume 8609, pp 366-375.
[13] Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. Journal of Journal of the American Medical Informatics Association, 2007: 14(3):304-11.
[14] Cruz Díaz NP, Maña López MJ, Mata Vázquez J, Pachón Álvarez V. A machine-learning approach to negation and speculation detection in clinical texts. Journal of the American Society for Information Science and Technology, 2012. 63(7): 1398-1410.
[15] Bosmans JM., Peremans L, Menni M, De Schepper AM, Duyck PO, Parizel PM. Insights Imaging. Structured reporting: if, why, when, how-and at what expense? Results of a focus group meeting of radiology professionals from eight countries, 2012: 3(3):295-302.

Address for correspondence

Viviana Cotik: vcotik@dc.uba.ar.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-639

Automated Learning of Temporal Expressions

Douglas Redd[a,b] MS, YiJun Shao[a,b] PhD, JingYang [a,b] BS, Guy Divita [a,b] MS, Qing Zeng-Treitler [a,b] PhD

[a] Salt Lake City Veteran Affairs Hospital, Salt Lake City, Utah, USA, [b] University of Utah, Salt Lake City, Utah, USA

Abstract

Clinical notes contain important temporal information that are critical for making clinical diagnosis and treatment as well as for retrospective analyses. Manually created regular expressions are commonly used for the extraction of temporal information; however, this can be a time consuming and brittle approach. We describe a novel algorithm for automatic learning of regular expressions in recognizing temporal expressions.

Five classes of temporal expressions are identified. Keywords specific to those classes are used to retrieve snippets of text representing the same keywords in context. Those snippets are used for Regular Expression Discovery Extraction (REDEx). These learned regular expressions are then evaluated using 10-fold cross validation. Precision and recall are very high, above 0.95 for most classes.

Keywords:

Electronic Medical Record, Machine Learning

Introduction

Temporal expressions are a special type of named entity. In clinical notes, temporal information is often critical to the interpretation of findings. The time order of events is an essential factor in assessing causation and in evaluating co-occurrence [1]. For example, an adverse reaction to a treatment can only be shown if the reaction occurred after the treatment. Additionally, drug-drug interactions can only be shown if the drugs were taken within the same time frame.. In prior studies that extracted temporal expressions, regular expressions are commonly used [2]. While regular expressions are powerful, they do typically need to be manually created based on chart reviews. This creates a challenge for maintenance and adaptation. In this paper, we describe the use of a simple and novel learning algorithm to discover temporal expressions. While the expressions we discovered are specific to the set of training data employed by this study, the algorithm is generalizable to other training datasets and to other types of expressions.

Background

The temporal information in medical records is important for both clinical decision support and general medical research. This information exists as both structured data and unstructured narrative data. Extraction of temporal information from structured data is trivial but is much harder from the narrative data [3]. For instance, temporal expression extraction was part of the task proposed by the 2012 i2b2 NLP for Clinical Data challenge [4].

Earlier studies have attempted to identify the temporal information using NLP techniques in a simplified form, e.g., "historical" vs. "current" [5,6]. More recent studies have shown that it is feasible to extract the exact temporal expressions from clinical narratives. For example, in one study [7] a system called Med-TTK was developed for detecting temporal expressions in medical narratives by extending an existing system called Temporal Awareness and Reasoning Systems for Question Interpretation Toolkit (TTK) [8]. TTK is an open-source software package developed for extracting temporal information in news articles. The Med-TTK system achieved an overall F-score of 0.85 on a set of 200 U.S. Department of Veterans Affairs (VA) clinical notes. Another study [9] combined rules and machine learning for the extraction of temporal expressions and achieved a micro F-score of 0.90 on a set of 320 clinical notes provided by the 2012 i2b2 challenge.

Methods

Corpus

We used a Pittsburgh EHR dataset (Pitts Data) for this study [10]. Pitts Data is composed of 100,866 de-identified clinical notes and was used in the Text Retrieval Conference (TREC) Medical Track [11], an annual event for encouraging the development of medical information retrieval techniques.

Keyword Identification

An initial set of temporal keywords were constructed from TIMEN, a temporal expression normalization tool [12]. We used 193 keywords that included time of a day, month, season, decade, and holidays. Sample keywords are shown in table 1.

Table 1 – Sample Temporal Expression Keywords

at	pm
after	midday
before	afternoon
between	noon
by	evening
during	overnight
following	midnight
for	night
from	pm
on	p m
since	pm
till	Seconds
to	second
until	second
within	minutes
while	minute
when	Hours
except	hour
in	days
a.m.	day

Snippet Extraction

The set of temporal keywords were used to extract a set of snippets from the document corpus. We used snippets to refer to the keyword and the context surrounding the keyword in the document. In this study, each snippet consisted of the matched keyword, the 10 words preceding the keyword in the document, and the 10 words following the keyword in the document. We progressively extracted all snippets containing the keywords from documents until we surpassed our target of 1,000 snippets for annotation. This resulted in an annotation set of 1,008 snippets from 13 documents.

Annotation

Human annotation using VTT [13] was performed to indicate the exact temporal expression in each snippet and the class of the temporal expressions (figure 1). The classes were DATE, TIME, DURATION, SET, and OTHER, with examples given in table 2. The temporal expression was marked in each snippet (referred to as the "labeled segment"), and it's temporal class assigned, by two independent annotators. To reach agreement, two rounds of annotation testing were performed on subsets of data. The kappa for the second round was 94%. Afterwards one annotator performed the annotation for the remaining documents.

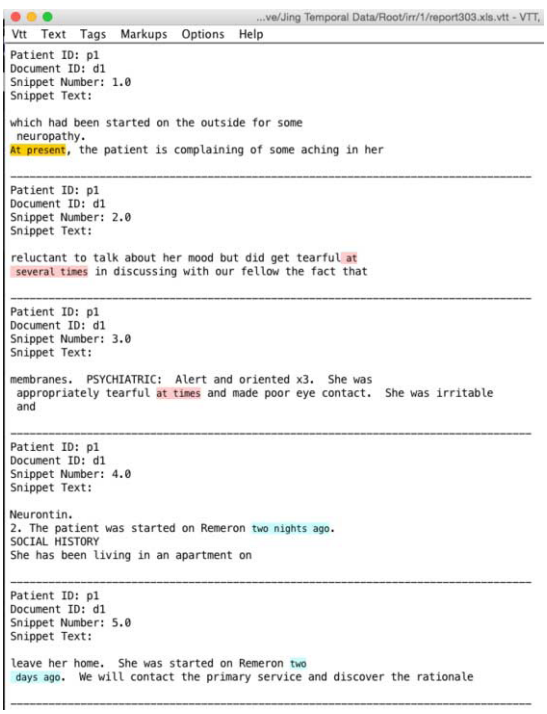


Figure 1 – Sample of VTT annotations.

Information Extraction

We trained a regular expression discovery Extraction (REDEx) algorithm on the annotated snippets. The REDEx algorithm is a novel process we have developed that automatically learns regular expressions that capture the value of the annotated labeled segment and the context surrounding it. We have used previous versions of REDEx for various value extraction tasks, including body weight [14]. We implemented REDEx in Java, making extensive use of the java.util.regex.* core libraries. It is open source and licensed under the Apache License, Version 2.

Table 2 – Temporal Expression Classes

Class	Description	Examples
DATE	Regarding a specific day.	Jan 14 2007
TIME	A specific time point.	14:04:28
DURATION	A period of time.	40 minutes
SET	A set of several temporal expressions.	Monday and Wednesday
OTHERS	Temporal expressions with vague resolution.	past

The current version of REDEx is represented as pseudo-code in Figure 3. Briefly, each annotated snippet was first split into 3 parts: the "labeled segment (LS)", which is the piece of text marked by the annotator; the "before labeled segment (BLS)", which is the text between the beginning of the snippet and the start of the LS; and "after labeled segment (ALS)", consisting of the text between the end of the LS and the end of the snippet (figure 2). Each BLS-LS-ALS triplet was then generalized to a regular expression by replacing all punctuation, whitespace, and digits with corresponding regular expressions: \p{Punct}, \s{1,50}, and \d+ respectively. The BLS-LS-ALS triplets were then iteratively generalized by successive rounds of trimming from the front of the BLS and the end of the ALS until one or more false positives were observed. We interpreted a false positive to be a case were REDEx predicted a value where there was none in the manual annotations. Duplicate triplets were removed, and then the triplets were combined into a single regular expression. The LS was marked as a capture group in order to retrieve the matched value. Sensitivity was then calculated for each regular expression by dividing the count of matched snippets by the total number of snippets.

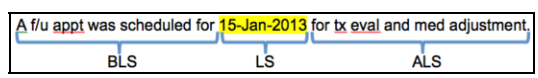


Figure 2 - Before Labeled Segment (BLS), Labeled Segment (LS), and After Labeled Segment (ALS) of a date expression in a phrase.

Evaluation

Evaluation of the regular expressions was performed using 10-fold cross validation, with final scores being the mean of those from each of the folds. Evaluation measures of precision, recall, F1-score, and accuracy were calculated for each class, along with the number of snippets in each class and the number of unique regular expressions generated by REDEx. We defined predictions as true positive (TP) when the regular expressions extracted a value from a snippet that matched the annotated value, false positive (FP) when the extracted value did not match the annotated value, false negative (FN) if the regular expressions failed to extract a value but there was an annotated value, and true negative (TN) if there was not an extracted value and there was no annotated value.

Results

1,008 snippets were extracted for annotation using 193 keywords from the corpus. These snippets were human annotated to identify the temporal expression spans and classified as DATE, TIME, DURATION, SET, or OTHER. Regular expressions and their sensitivities were then automatically generated using the RED extraction algorithm for each class. Samples of the resulting regular expressions are shown in table 3.

Evaluation measures of precision, recall, F1-score, accuracy, number of snippets in each class, and the number of unique regular expressions generated by REDEx are presented in table 4. Evaluation metrics were very high in most cases. Date, Time, and Set classes were all ≥ 0.97 for precision. Recall was ≥ 0.96 for Date, Time, Duration, and Other. The only measures below 0.90 were for the Set class, where recall and accuracy were 0.83 and the sample size was very small at 18.

Table 3 – Examples of REDEx Regular Expressions

Class	Regular Expression
DATE	(ten\s{1,50}days\s{1,50}ago) \p{Punct}DATE\p{Punct}(Aug\s{1,50}\d+\s{1,50}\d+)
TIME	(\d+\p{Punct}\d+\p{Punct}\d+\s{1,50}AM) \s{1,50}\d+\s{1,50}\d+\p{Punct}\s{1,50}(\d+\p{Punct}\d+\p{Punct}\d+)\s{1,50}T
DURATION	\s{1,50}(another\s{1,50}four\s{1,50}weeks) \s{1,50}\S{1,6}\s{1,50}\S{1,6}\s{1,50}\S{1,10}\p{Punct}\s{1,50}on\s{1,50}a\s{1,50}\S{1,5}\s{1,50}\S{1,5}\s{1,50}\S{1,5}\p{Punct}\s{1,50}over\s{1,50}(the\s{1,50}past\s{1,50}few\s{1,50}months)
SET	(every\s{1,50}\d+\s{1,50}hours) (per\s{1,50}hour)
OTHERS	(recently)\s{1,50}discharged\s{1,50}from\s{1,50}the\s{1,50}hospital\s{1,50}and\s{1,50}\S{1,10}\s{1,50} \s{1,50}\S{1,4}\s{1,50}minimally\s{1,50}invasive\s{1,50}esophagectomy\s{1,50}with\s{1,50}reanastomosis\s{1,50}known\s{1,50}to\s{1,50}our\s{1,50}service\s{1,50}for\s{1,50}(recent)

Table 4 – Evaluation Metrics

Measure	Date	Time	Duration	Set	Other
Precision	0.97	0.98	0.93	1.00	0.93
Recall	0.97	0.97	0.96	0.83	0.96
F1-score	0.97	0.98	0.95	0.91	0.95
Accuracy	0.94	0.95	0.90	0.83	0.90
# Snippets	493	289	169	18	39
# Reg Ex	128	54	41	7	17

```
BLS = Before Labeled Segment
LS = Labeled Segment
ALS = After Labeled Segment

PS = Positive Text Snippet
NS = Negative Text Snippet

Regular Expression Discovery (PS, NS)

/*Initialize Result*/
RS = { }

/*For each positive instances*/
For each p in PS
        /*Transform a text string into regular expression
        by replacing punctuations, white spaces and dig-
        its*/
        P' = Generalize (p);
        /*Split each expression into 3 segments*/
        (bls', ls', als') = Split (p');
End For

/*Combine all labeled patterns*/
ls_exp = Combine (LS');

/*For each positive instance*/
For each p
        /*Test the regular expression using the negative
        instances; if an expression matches any negative
        instances, it is discarded. */
        While match (p', NS) == False
                /*Trim the segments before or after the
                labeled segments*/
                p'' = trim (bls', ls', als');
        End While

        /*Add the shortest regular expression that does
        NOT match any snippets in the negative sam-
        ple*/
        RS = RS + p';
End For
```

Figure 3 - Pseudo-code describing the RED Extraction algorithm

Discussion

This study has demonstrated the feasibility of automatically discovering temporal expressions. Given the moderate amount of training data, we were able to achieve a very high level of sensitivity and specificity. The machine-generated expressions are humanly readable, though often not as succinct as the expression a senior programmer would have written. It is also worth noting that, the REDEx algorithm does not require seed patterns to begin with.

The REDEx algorithm can be applied to other use cases as well. We have, for instance, used REDEx to extract weight value and unit, with the same level of sensitivity and specificity as for temporal extraction [14]. When using REDEx, we trade the time required for writing and testing regular expression with the time required for annotation. In the use cases of temporal information and weight value/unit, we found the trade-off to be a beneficial. In these two cases,

research assistants and researchers who do not have programming knowledge did the annotations very quickly (100 to 150 annotations/hour after the inter rater agreement is established). The amount of time it takes REDEx to generate expressions was trivial comparing to the manual generation of expressions, thus REDEx can potentially incorporate many more examples. The development of REDEx itself did take time. It, however, can be re-used. For the purpose of future maintenance, we felt that it would be much easier to create additional annotations than to manually revise the expressions.

There are many machine learning algorithms for text classification, including for context classification. There are considerably fewer learning algorithms that we can readily use for the discovery of specific (regular) patterns. With REDEx, we can learn to recognize sequential patterns with numbers, symbols and letters. Although developed and evaluated using the English alphabet, this should be directly applicable to other Latin alphabets using Arabic numerals.

The Pitts dataset used for training and testing is relatively uniform. Our concern is that it does not contain sufficient variations of the temporal expressions. We plan to extend the learning and testing to other clinical notes and create a larger set of temporal expressions. In addition, we plan to evaluate the REDEx algorithm in comparison to manually created regular expressions and possibly the Med-TTK. Other future experiments are to evaluate with other corpora, and to determine the accuracy of regular expressions discovered in one corpus when applied to a different corpus.

We performed the progressive snippet selection method which did not differentiate between the temporal classes at the time of selection. This resulted in small sample sizes for the Set and Other classes. In future studies the selection method can be improved to provide better representation of all classes. This can be done using a similar progressive selection method, but counting the snippets separately for each class. Or by extracting all snippets, stratifying them by class, then randomly sampling an equal number from each stratum. This would have the added advantage of providing class cardinality across the corpus as well as better representation of samples.

The temporal expressions discovered in this study have been incorporated into a v3NLP Framework NLP module [15] to look up temporal text from clinical notes. Additional classes of temporal expressions, in particular intervals, are planned as additions. When extended and tested on other datasets, the NLP module will be released as open source software.

Conclusion

We have developed and tested a simple and novel REDEx algorithm for the discovery of temporal expressions, with good sensitivity and specificity. The REDEx can be applied to other types of information extraction tasks, because it does not contain any built-in information about temporal data.

Acknowledgements

This work was funded by VA grants: NLP Ecosystem CRE 12-315, CHIR HIR 08-374 and VINCI HIR-08-204.

References

[1] Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965 May;58(5):295-300. PubMed PMID: 14283879.

[2] Babbar R, Singh N. Clustering based approach to learning regular expressions over large alphabet for noisy unstructured text. Proceedings of the fourth workshop on Analytics for noisy unstructured text data; Toronto, ON, Canada. 1871848: ACM; 2010. p. 43-50.

[3] Zhou L, Hripcsak G. Temporal reasoning with medical data—A review with emphasis on medical natural language processing. Journal of Biomedical Informatics. 2007 4//;40(2):183-202.

[4] Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. Journal of the American Medical Informatics Association. 2013;20(5):806-13.

[5] Chapman WW, Chu D, Dowling JN. ConText: an algorithm for identifying contextual features from clinical text. Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing; Prague, Czech Republic. 1572408: Association for Computational Linguistics; 2007. p. 81-8.

[6] Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. Journal of Biomedical Informatics. 2009 10//;42(5):839-51.

[7] Reeves RM, Ong FR, Matheny ME, Denny JC, Aronsky D, Gobbel GT, et al. Detecting temporal expressions in medical narratives. International Journal of Medical Informatics. 2013 2//;82(2):118-27.

[8] Verhagen M, Mani I, Sauri R, Knippen R, Jang SB, Littman J, et al. Automating temporal annotation with TARSQI. Proceedings of the ACL 2005 on Interactive poster and demonstration sessions; Ann Arbor, Michigan. 1225774: Association for Computational Linguistics; 2005. p. 81-4.

[9] Kovačević A, Dehghan A, Filannino M, Keane JA, Nenadic G. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. Journal of the American Medical Informatics Association. 2013;20(5):859-66.

[10] Voorhees EM, Hersh W. Overview of the TREC 2012 Medical Records Track. Available from: http://trec.nist.gov/pubs/trec21/papers/MED12OVERVIEW.pdf.

[11] Edinger T, Cohen AM, Bedrick S, Ambert K, Hersh W. Barriers to Retrieving Patient Information from Electronic Health Record Data: Failure Analysis from the TREC Medical Records Track. AMIA Annual Symposium Proceedings. 2012 11/03;2012:180-8. PubMed PMID: PMC3540501.

[12] Llorens H, Derczynski L, Gaizauskas R, Saquete E, editors. TIMEN: An Open Temporal Expression Normalisation Resource 2012: European Language Resources Association (ELRA).

[13] Visual Tagging Tool: United States National Library of Medicine; 2010. VTT]. Available from: http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/vtt/current/web/index.html.

[14] Murtaugh MA, Gibson BS, Redd D, Zeng-Treitler Q. Regular expression-based learning to extract bodyweight values from clinical notes. Journal of Biomedical Informatics. (0).

[15] Divita G, Zeng-Treitler Q, Gundlapalli AV, Duvall S, Nebeker J, Samore MH. Sophia: A Expedient UMLS Concept Extraction Annotator. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2014:467-76.

Address for correspondence

Qing Zeng-Treitler <q.t.zeng@utah.edu>

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-643

Identifying Diseases, Drugs, and Symptoms in Twitter

Antonio Jimeno-Yepes[*a,b], Andrew MacKinlay[*a,b], Bo Han[a], Qiang Chen[a]

[a] *Melbourne Research Lab, IBM Research, Victoria, Australia*
[b] *Dept. of Computing and Information Systems, University of Melbourne, Australia*

Abstract

Social media sites, such as Twitter, are a rich source of many kinds of information, including health-related information. Accurate detection of entities such as diseases, drugs, and symptoms could be used for biosurveillance (e.g. monitoring of flu) and identification of adverse drug events. However, a critical assessment of performance of current text mining technology on Twitter has not been done yet in the medical domain. Here, we study the development of a Twitter data set annotated with relevant medical entities which we have publicly released. The manual annotation results show that it is possible to perform high-quality annotation despite of the complexity of medical terminology and the lack of context in a tweet. Furthermore, we have evaluated the capability of state-of-the-art approaches to reproduce the annotations in the data set. The best methods achieve F-scores of 55–66%. The data analysis and the preliminary results provide valuable insights on identifying medical entities in Twitter for various applications.

Keywords:
Social media, data set generation, medical entity recognition.

Introduction

The volume of data on social media sites, such as Twitter, is so vast that it would almost be surprising if it did not contain useful medical information. If we could successfully mine even a small percentage of this, there would be many potential uses, including biosurveillance (e.g. monitoring of seasonal flu) and identification of potential adverse drug events. Previous work exists on biosurveillance based on emergency department notes [1], news data [2], and search data [3], while additional work exists on the detection of adverse effects extracted from forum data [4-6] and Wikipedia [7].

While there has been some work on medical text mining in social media (i.e. identification of relevant tweets for adverse drug events [8]), a critical assessment of performance of current text mining technology has not been performed. It has already been established that Twitter itself presents unique challenges for text mining in the open domain [9,10].

In this work, we have developed an annotated data set from Twitter feeds that can be used to train and evaluate methods to recognise mentions of diseases, symptoms, pharmacologic substances in social media, and particularly microblogs. Furthermore, we have evaluated the performance of existing state-of-the-art entity recognition approaches on this data set. Overall, methods based on conditional random fields allow high precision entity recognition, while additional work is required to improve the recall.

[*] The first and second authors contributed equally to this paper

Materials and Methods

This section presents the development of the Twitter data set and the annotation process, as well as the methods used to automatically reproduce the annotation.

Data Collection And Filtering

We obtained our data using Twitter Stream API from 13/05/2014 to 28/05/2014, collecting 43 million tweets in total. An inspection of random samples indicated that most tweets do not contain medical entities. We pre-filtered tweets using a list of medical terms. We considered three types of entities: *diseases*, *symptoms,* and *pharmacologic substances* to match the particular entities we were targeting for annotation. The list of these biomedical entities comes from the Unified Medical Language System (UMLS) [11] Metathesaurus, which integrates over 100 biomedical terminologies and ontologies. The concepts in the Metathesaurus are assigned one or more semantic types from the UMLS Semantic Network[i] [12]. We downloaded the UMLS version UMLS2014AA and used the default installation. Within this network, we selected the following semantic type mappings:

- T047 (*Diseases or Syndrome*) for diseases,
- T184 (*Sign or Symptom*) for symptoms,
- T121 (*Pharmacologic Substance*) for pharmacologic substances.

From the Metathesaurus, we only retained concepts linked to one of the preceding semantic types, and then extracted the union of terms corresponding to these concepts.[ii] Furthermore, only terms in English and non-obsolete entries were kept. Table 1 shows statistics for each entity type. Disease and pharmacologic substances contain a large number of concepts and terms, while the list of symptoms is comparatively small.

Table 1 – Statistics of the concepts and terms extracted from the UMLS for the three entity types.

Entity	Concept-term pairs	Unique terms	Unique concepts
Disease	201,916	201,013	46,993
Pharm. Subs.	279,261	278,390	120,330
Symptom	19,927	16,865	3,850

[i] UMLS Semantic Network site: http://semanticnetwork.nlm.nih.gov
[ii] We joined the relevant Metathesaurus tables ('MRCONSO' and 'MRSTY') to determine this information.

We filtered the Twitter data to keep the volume manageable. Only tweets with two or more entity types (e.g. a *symptom* and *disease*) were kept, ensuring the filtering is fairly precise (while potentially reducing recall). The lists of UMLS terms extracted, as described above, contain very frequent terms that are primarily used in a non-medical sense, such as *said* and *water*. To avoid a large number of false positives in the filtering stage, we used the frequency of UMLS terms in filtered tweets from the previous step and ranked them in decreasing order of frequency. We then manually removed common terms with a non-medical primary sense from the top 200 terms in each type.

We further removed duplicates in the filtered tweets and removed non-English tweets using LANGID.PY [13]. This filtering pipeline ultimately yielded 11,647 tweets. To investigate whether our filtering (particularly our requirement of the presence of two entity types) erroneously excluded relevant tweets, we examined 1,000 randomly selected tweets. In this sample, we found no relevant tweets—that is, tweets containing biomedical entities of interest—which would have been excluded, suggesting that this pre-filtering methodology is fairly reliable for including all possibly relevant data.

Data Set Annotation Procedure

Four annotators with no medical training annotated the data set with entities from the three semantic types, using BRAT [14]. We created the guidelines iteratively as described here in chronological order.

We prepared an initial set of guidelines based on manual examination of a small subset of the data which was not used for later annotation. We then had all four annotators use these initial guidelines to annotate the same set of 100 tweets as a calibration set.

We checked the inter-annotator agreement on the calibration set and found that it was too low to be appropriate for what we would like to consider a high-quality data set. On the basis of discussions between the annotators, we then refined the guidelines to attempt to resolve frequently observed points of ambiguity in the calibration set. A subsequent round of annotation with the new guidelines improved the inter-annotator agreement, but still not to a level we considered acceptable. At this point we moved to a system of having all tweets annotated by two annotators, then merging the annotations and having the annotators resolve all disagreements in discussion.

This double annotation methodology led to a higher-quality data set with very respectable figures for inter-annotator agreement as shown below. The cost, of course, was that we were able to obtain less data per hour of annotation time. In practice we found the merging and discussion process was generally fast (roughly 20% of the annotation time itself), meaning that overall annotation efficiency was reduced by a factor of around 2.4 by the double annotation methodology. However, for future iterations of the data set, we will leave open the option of augmenting this high quality data set with a lower-quality data set annotated only by a single annotator (possibly assisted by automatic pre-annotation).

Since we settled on double-annotation, the relevant comparison for inter-annotator agreement is to calculate agreement on a subset of the data, which has been annotated twice in the manner previously described. That is, the four annotators were grouped into two pairs, and each pair annotated and merged the same subset of 100 tweets using the methodology described above. We then calculated agreement figures between the two sets of annotations, obtained using BratEval[iii] [15].

Table 2 – Inter annotator agreement for each one of the entity types.

Entity	Precision	Recall	F1
Disease	0.8400	0.8750	0.8571
Pharm. Subs.	0.9500	0.8261	0.8837
Symptom	0.8246	0.8393	0.8319

Inter annotator agreement, as shown in Table 2, is reassuringly high, particularly for such a potentially ambiguous task, and displayed similar levels of agreement to other biomedical annotation tasks [16]. Most of the disagreements were terms inadvertently missed by the annotators, and in a few cases the words were arguably non-medical terms, which can sometimes be difficult to distinguish. For instance, terms like *chill* or *weak* often carry a non-medical meaning. Other disagreements included different interpretations of the subtleties of the guidelines, such as whether *pill* was sufficiently specific to be included.

Annotation Guidelines

After the two rounds of annotation calibration, we settled on a final set of annotation guidelines. These stipulated that we are interested in annotating three kinds of entities: pharmacologic substances, diseases, and symptoms. In addition to traditional entities, which correspond to noun phrases, we also broadened the scope of the annotation to allow for short phrases headed by verbs (such as *I **coughed** all morning*) and adjectives (such as *felt **light-headed***), which indicate diseases or symptoms. If the part-of-speech of the head-word of an annotated item is not a noun, the annotation was marked as an adjective or verb as appropriate in a separate attribute.

We also found that in many cases a concept being mentioned, which may have looked superficially medical, was unlikely to truly refer to a medical concept. In particular, mentions may be metaphorical, figurative, or purely humorous, and in these cases annotators were instructed to apply the 'figurative' attribute. There was also another slightly distinct class of items, one in which the terms have an informal and non-clinical meaning in addition to the clinical one. If the non-clinical meaning is clearly being used, annotators were to apply the 'non-medical' attribute, such as in ***depressed** about my exam results*.

The guidelines also instruct annotators to annotate the most specific entity possible (e.g. *codeine syrup* rather than *syrup*) and to include as many tokens as possible as long as they are part of a fixed expression referring to a particular kind of entity (e.g. *disgusting* would not be part of the entity in *disgusting codeine syrup*). It was also specified that entities which do not distinguish anything more specific than the base entity category should not be annotated, as the very general information in these is unlikely to be specific enough to be useful in downstream applications. In addition, it was also permitted to have overlapping annotations, so while *pain medication* would most sensibly be annotated as a pharmacologic substance, the token *pain* within it should also be annotated as a symptom.

Correctly identifying medical concept mentions and categorising them is sometimes a difficult task for annotators without formal medical training as there is a large terminology space. So it may be difficult to determine whether a particular token refers to a valid concept (e.g. Should *prune juice* be

[iii] BratEval site: https://bitbucket.org/nicta_biomed/brateval

considered a pharmacologic substance? Is *dextromethorphan* a real drug name?), or which of two categories it should refer to. In particular, the division between the disease and symptom categories can be very uncertain in many cases. For example, it may not be clear initially whether *amnesia* is a disease or a symptom. So, the annotators were advised to refer to the UMLS in cases of uncertainty, essentially using the UMLS as a substitute for in-depth domain knowledge. In particular, the UMLS semantic type is important, so generally the semantic types should obey the same mapping as described in 'Data collection and filtering' (for example, *disease* entities should have semantic type 'T047').

Even if the UMLS was not the perfect resource for making these decisions, it is in widespread usage (incorporating many standardised terminologies), and at least provides a common basis for decision-making, ensuring consistency of the annotations. In some cases, the context may make it clear that a strict interpretation of UMLS semantic type, as described above, is not appropriate; in these cases, annotators were free to apply their own more appropriate categorisation instead.

It is of course difficult to codify every possible annotation decision in a static set of guidelines. Inevitably, certain borderline cases had to be decided on the basis of the annotators' intuitions in ways that are difficult to encode specifically. However, the procedure of double annotation that we adopted and the high inter-annotator agreement we achieved at least suggests that the annotations are reasonably internally consistent, and thus presumably repeatable.

Data Set Statistics

The final data set contains 1,300 annotated tweets in 13 files with 100 tweets each. As shown in the following table, *symptom* is the most frequent type with no significant difference between disease and pharmacologic substance. We can see that entities are typically composed of a single token for *symptoms*, while *diseases* and *pharmacologic substances* more frequently span multiple tokens.

Table 3 – Statistics for the entities in the data set. Entities annotated as non-medical have not been considered.

Entity	No Entities	Avg. length	Avg. Tokens
Disease	253	10.40 ± 5.83	1.41 ± 0.62
Pharm. Subs.	233	9.83 ± 4.35	1.39 ± 0.58
Symptom	764	6.66 ± 2.96	1.13 ± 0.40

Table 4 shows the number of non-medical and figurative terms annotated. The number of pharmacologic substances and symptoms that are not medically related are significantly larger; it seems these terms are more often used informally, rather than with their medical definition. In addition, almost 10% of the symptom mentions are used figuratively.

Table 4 – Number of non-medical and figurative entities.

Entity	Non-medical	Figurative	Medical Figurative
Disease	1 (0.40%)	5 (1.98%)	4 (1.58%)
Pharm. Subs.	20 (8.58%)	2 (0.86%)	2 (0.86%)
Symptom	122 (9.65%)	124 (9.81%)	44 (3.48%)

Table 5 – Part-of-speech of the annotated entities. Entities annotated as non-medical have not been considered.

Entity	Noun	Adjective	Verb
Disease	246 (97.2%)	7 (2.8%)	0 (0.0%)
Pharm. Subs.	233 (100.0%)	0 (0.0%)	0 (0.0%)
Symptom	454 (75.5%)	262 (20.7%)	48 (3.8%)

Almost all entities annotated are nouns as noted in Table 5. It is logical that pharmacologic substances are nouns. There are only a few mentions of diseases that appear as adjectives (e.g. *blind*, *obese,* or *overweight*). For symptoms, over 20% appear as adjectives (e.g. *breathless*, *hungry,* or *sick*) and a smaller quantity appear as verbs (e.g. *fainted*, *coughing,* or *shaking*).

Table 6 shows the top 10 terms by frequency per entity type. There is a large number of individual posting on Twitter and we can identify multiple topics in our data set, which shows the possibilities of exploiting Twitter data and the complexity of extracting relevant signals from it. Some of these terms denote concerns about diseases that affect a large part of the population (e.g. *diabetes*) but also highlight recent breakthroughs in medicine (e.g. a new *malaria vaccine*). We also find mentions of recreational drugs (such as *marijuana* and *cannabis*), which are not related to any specific news item. In addition, there are mentions like *heroin* linked to news when the tweet was posted (e.g. *NYPD officers to carry heroin overdose antidotes*). Furthermore, we find a large number of symptoms that are not linked to any specific disease, which could be monitored as signals for biosurveillance.

Table 6 – Top most frequent terms per entity type.

Disease		Phar. Sub.		Symptom	
diabetes	14	marijuana	12	tired	136
heart disease	10	cannabis	12	pain	93
stroke	10	alcohol	11	hungry	61
cold	8	heroin	8	stress	50
asthma	6	pain meds	7	headache	36
malaria	6	vitamin c	4	sore	15
allergy	4	chill pill	4	sick	14
migraine	4	malaria vaccine	4	cough	13
aids	4	caffeine	4	exhausted	12
obesity	4	calcium	4	hangover	11

Named Entity Recognition

After annotating a data set, we investigated how applicable standard approaches to named entity recognition (NER) would be to this data. This indicates how unique the data is, and helps us to predict how difficult it will be to reliably reproduce these annotations automatically on unseen data, which is our ultimate goal. Three methods were used to annotate the data set with the three entity types.

The first uses MetaMap [16], which is developed at the US National Library of Medicine that maps spans of text to UMLS Metathesaurus concepts and is considered state-of-the-art for this task. It uses parsing to identify spans of text in which entities could appear and then smart dictionary matching to identify the concepts [16]. We used MetaMap 2013, with its default configuration, to perform experiments with and without word sense disambiguation (WSD) [18]. MetaMap output was filtered to keep only concepts belonging to the three semantic types under consideration, and the output was converted to BRAT standoff format for evaluation.

In addition to MetaMap, we considered two systems based on machine learning (ML). Firstly, we created a custom NER tagger for this data set, called 'Micromed' (since it tags medical concepts in microblog posts), to provide an in-house solution. It uses conditional random fields (CRF) [19] as its underlying machine learning algorithm. CRFs are frequently used in state-of-the-art named entity recognizers, including those in the biomedical domain. For its CRF implementation, Micromed uses CRFSuite [20]. The features were derived

*Table 7 – Results of all automatic annotators over all entity types. Statistical significance at p < 0.01: * vs MetaMap; † vs SNER.*

		Disease			Pharm. Substance			Symptom		
	Method	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**
Exact Match	MetaMap +WSD	0.4123	0.5850	0.4837	0.2264	0.5150	0.3145	0.5337	0.5604	0.5467
	MetaMap	0.3437	0.7876	0.4786	0.2225	0.7785	0.3460	0.4644	0.7635	0.5775
	Stanford NER	0.7917	0.3071	0.4312	0.8952	0.3565	0.4946*	0.7526	0.5763	0.6509*
	Micromed	0.7987	0.5020	0.6165*†	0.8142	0.3948	0.5318*	0.7193	0.6028	0.6559*
	Micromed +Meta	0.8049	0.5217	**0.6331***†	0.8205	0.4120	**0.5486***	0.7220	0.6041	**0.6578***
Partial Match	MetaMap +WSD	0.4457	0.6299	0.5220	0.2642	0.5957	0.3660	0.4424	0.6150	0.5146
	MetaMap	0.3437	0.7876	0.4786	0.2225	0.7785	0.3460	0.4644	0.7603	0.5766
	Stanford NER	0.7917	0.3071	0.4312	0.8952	0.3565	0.4946*	0.7603	0.6013	0.6696*
	Micromed	0.7987	0.5020	0.6165*†	0.8142	0.3948	0.5318*	0.7439	0.6234	0.6783*
	Micromed +Meta	0.8049	0.5217	**0.6331***†	0.7220	0.6041	**0.6578***	0.7450	0.6234	**0.6788**

from those commonly used in NER, and tuned somewhat to this particular task:

- Part-of-speech tag and relative position in a context window of three tokens each side
- Token surface form and relative position, in a context window of two tokens each side
- Token prefix and suffix character N-grams of all lengths up to eight
- Whether the token appears in a list of synonyms for concepts with the appropriate semantic type extracted from UMLS
- (In some configurations) whether MetaMap annotates the token as a concept with that semantic type

Since the goal was to evaluate how similar the task was to standard NER, we did not add particularly radical features. For tokenising and POS-tagging, we used TweetNLP [21]. We trained a CRF model for each category (*diseases*, *pharmacologic substances,* and *symptoms*), treating them as distinct annotation tasks. The output of the CRF engine was converted to BRAT standoff[iv] format for evaluation.

Secondly, we used the Stanford NER tagger (SNER) [22] as another strong baseline to study how difficult the task is for existing NER tools and help identify effective features. This also underlyingly uses a CRF. We reformatted annotations into SNER format and applied the limited default features to train taggers for each of the three categories, which included character n–grams and word tokens in fixed context windows (the primary feature difference from Micromed being the absence of custom lookup lexicons).

Results

To compare the performance of the annotation methods, we used two different evaluation profiles. Exact match requires a given entity from the classifier output to have the same start and end span as the reference entity to be considered a match. Partial match considers entities as matching if there is any overlap at all between the entity produced by the classifier and the gold standard. After counting matches, we calculated precision, recall, and F1 in the usual way. Entity annotations marked as non-medical were ignored in training and testing. Statistical significance of the results shown in Table 7 was computed using a two-sample *t*-test with randomization over the cross-validation folds.

Machine-learning methods were trained and evaluated using 13-fold cross validation, based on 13 sections of the data set, each containing 100 tweets. That is, at each of the 13 iterations, 1200 tweets (12 sections) were used for training, while the remaining 100 tweets were used for evaluation. MetaMap, however, was simply applied to the whole data set, so the results are comparable. We evaluated Micromed, both with and without features based on MetaMap (the former case is denoted "+Meta"), to evaluate how well it could perform without relying on an external tool with a significant overhead. Results of the overall methods are presented in Table 7.

MetaMap results without WSD have much higher recall, but poorer precision. Entities missed include terms that are not in the UMLS (e.g. *painkillers addiction*) or terms in the UMLS that are not in the categories of interest (e.g. *cold* as *symptom*). False positives include non-specific terms annotated by MetaMap (e.g. *drugs*), which are excluded by the annotation guidelines, terms bearing a non-medical meaning (e.g. *I'm sick and tired of negativity),* and WSD mistakes (e,g. *cannabis* was annotated as *plant* instead of *substance)*.

The ML methods usually outperforms MetaMap. SNER's accuracy is lower over *Disease* but substantially higher over *Pharmacologic Substances* and *Symptom*. Micromed has higher F-score again, with statistically significant increases over SNER except for *Symptom.* SNER has lower recall since it lacks the implicit domain knowledge from the UMLS-derived features of Micromed and Micromed+Meta (including MetaMap). We know, from Table 1, that the *Symptom* category has a smaller vocabulary; thus, relevant information can be learned from the training data alone. Moreover, some symptom entities, which SNER detects, are missed by Micromed; in many cases, these are not in the UMLS (e.g. *magnesium-deficient*). The increase in recall of partial match compared to exact match is not especially significant, except for *Symptom*.

Discussion

The data set was annotated with high inter-annotator agreement. Extra considerations were required for Twitter compared to biomedical literature, e.g., the *Figurative* attribute and extending the concept of entities beyond nouns, which might make traditional NER approaches perform poorly.

MetaMap was not trained on this data set as the other two methods were, but still shows a competitive performance with higher recall, while Micromed and SNER are generally more precision-biased and have a higher F-score overall. The

[iv] Brat standoff format: http://brat.nlplab.org/standoff.html

difference in performance of the methods is scientifically interesting in what it tells us about the nature of the data set. MetaMap is widely used for medical NER due to the respectable performance it achieves over research articles and clinical text. MetaMap has not been tuned for Twitter data and was outperformed in this work by ML-based classifiers, which could be effectively trained on a relatively small in-domain corpus. SNER even lacked in-domain knowledge, while Micromed used mostly standard NER features. This difference between MetaMap and ML methods suggests that the data set here has different characteristics to NER tagging in other domains.

The difference may also reflect different design considerations to be considered to tune MetaMap to work with Twitter data. In our work, the current performance of Micromed, as well as its lower computational cost[v] are likely to be advantageous in processing large volumes of social media data. The computational cost must, of course, be low if we hope to process a significant fraction of a high-volume data stream in real time. In addition, the smaller number of false positives generated by a higher precision system should lead to greater acceptance by users of the system output. Since the volume of data available in social media is relatively large with a high degree of redundancy with respect to overall trends, slightly lower recall is less of a concern in these applications.

Conclusion

In this work, we have presented the development of a Twitter set annotated with medical entities, showing that it is possible to perform high-quality annotation despite the complexity of medical terminology and the lack of context in a tweet. We have made the dataset publicly available[vi] to encourage further research. Furthermore, we have evaluated the capability of some state-of-the-art approaches to reproduce the manual annotations. The approaches demonstrate reasonable accuracy (with interesting variations between the methods), although further work is needed to identify additional features that might improve the performance of the annotators. We have focused on creating a data set and evaluating state-of-the-art annotators. We plan to use these annotators to process a live data stream from Twitter or some other source for biosurveillance and detecting adverse drug reactions.

References

[1] Espino, JU, Wagner MM, Tsui FC, et al. (2004). The RODS Open Source Project: removing a barrier to syndromic surveillance. Medinfo 2004: 1192-6.

[2] Collier N, et al. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 2008: 24(24): 2940-2941.

[3] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2008: 457(7232): 1012-1014.

[4] Segura-Bedmar I, de la Pena S, and Martínez P. Extracting drug indications and adverse drug reactions from Spanish health social media. ACL 2014, 98

[5] Metke-Jimenez A, Karimi S, and Paris C. Evaluation of text-processing algorithms for adverse drug event extraction from social media. Proceedings of the first international workshop on Social media retrieval and analysis. ACM, 2014.

[6] Cameron D, et al. "PREDOSE: A semantic web platform for drug abuse epidemiology using social media." Journal of Biomedical Informatics 2013: 46(6): 985-997.

[7] Generous N, et al. "Global disease monitoring and forecasting with Wikipedia." PLoS Computational Biology 2014: 10(11): e1003892.

[8] Ginn R, Pimpalkhute P, Nikfarjam A, et al.. Mining Twitter for Adverse Drug Reaction Mentions: a Corpus and Classification Benchmark, BioTxtM, LREC, 2014

[9] Baldwin T, Cook P, Lui M, et al. (2013). How noisy social media text, how diffrnt social media sources. In Proceedings of IJCNLP 2013.

[10] Han B, and Baldwin T (2011, June). Lexical normalisation of short text messages: makn sens a# twitter. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 368-378)

[11] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Research 2004: 32 (Suppl 1): D267-D270.

[12] McCray, AT. An upper-level ontology for the biomedical domain. Comparative and Functional Genomics 2003: 4(1): 80-84.

[13] Lui M, and Baldwin T. langid. py: An off-the-shelf language identification tool. Proceedings of the ACL 2012 System Demonstrations.

[14] Stenetorp P, Pyysalo S, Topić G, et al. brat: a Web-based Tool for NLP-Assisted Text Annotation. in *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France, 2012.

[15] Verspoor K, Jimeno-Yepes AJ, Cavedon L., et al. (2013). Annotating the biomedical literature for the human variome. Database 2013: bat019.

[16] Doğan, RI, Leaman R, and Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. Journal of Biomedical Informatics 2014: 47: 1-10.

[17] Aronson AR, and Lang FM. An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association 2010: 17(3), 229-236.

[18] Jimeno-Yepes A, and Aronson AR. Integration of UMLS and MEDLINE in unsupervised word sense disambiguation. AAAI Fall Symposium: Information Retrieval and Knowledge Discovery in Biomedical Text. 2012.

[19] Lafferty J, McCallum A, and Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. (2001).

[20] Naoaki O. CRFsuite: a fast implementation of conditional random fields (CRFs). URL http://www. chokkan. org/software/crfsuite (2007).

[21] Owoputi O, O'Connor B, Dyer C, Gimpel K, Schneider N, and Smith NA. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In Proceedings of NAACL 2013.

[22] Finkel JR, Grenager T, and Manning C. Incorporating non-local information into information extraction systems by gibbs sampling. Proceedings of ACL 2005.

Address for correspondence

Antonio Jimeno Yepes, IBM Research, Lvl 5, 204 Lygon street, Carlton, VIC, 3053, Australia – antonio.jimeno@au1.ibm.com

[v] Single-threaded processing time with MetaMap for this entire data set was 396 CPU-seconds, while Micromed over the same data took 19 CPU-seconds. However, it should be noted that MetaMap provides richer information than Micromed.

[vi] At https://github.com/IBMMRL/medinfo2015

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-648

Improving Preventive Healthcare with an User-centric Mobile Tele-monitoring Model

Valter Roesler[a,b], Alécio P. D. Binotto[b,c], Cirano Iochpe[a,b], Euro Bruno Palomba, MD[d] , Luiz A. P. Tizatto, MD[e]

[a]UFRGS – Federal University of Rio Grande do Sul, Brazil
[b]i9access Technology Ltda., Brazil
[c]IBM Research, Brazil
[d]Global Care Ltda.
[e]Unit Care Medical Services Ltda., Brazil

Abstract

Chronic diseases are an important field to tackle due to increasing healthcare risk factors, including population nutritional habits, lack of physical exercise, and population aging. Diabetes mellitus, hypertension, and obesity currently affect millions of people, and this statistic grows every year and is responsible for numerous deaths everyday. Many of those deaths could be delayed by following a steady monitoring strategy over such a population, which would prevent vital signs from reaching critical stages and providing knowledge for these patients about their health. This paper introduces Mobilicare, a mobile health promotion system designed to: (i) monitor remotely a patient´s vital signs in real time; (ii) support a health service in a Healthcare Center; and (iii) allow self-awareness of the disease and improve motivation. Our approach was applied to two distinct chronic patient management programs. The results showed the commitment of elder patients and the contribution of Mobilicare to the maintenance of a patient´s health stability.

Keywords:

Chronic disease management; eHealth; mHealth; cloud computing; tele-vital signs; tele-homecare.

Introduction

Access to medical care is sometimes difficult for citizens living in distant and underserved areas. However, even citizens living in large cities in developed countries may find it hard to reach medical services. Not being able to access primary medical care when illness, injury, or potential health problems arise may lead to a late diagnosis, delay in treatment, and possible future consequences. The problem gets even more serious in developing countries, such as the BRICS countries (Brazil, Russia, India, China, and South Africa).

With the goal of eliminating distance barriers, telemedicine is a multidisciplinary research area that integrates not only several computer science fields, but also medicine and healthcare. Its aim is to improve health services and quality of life, as well as provide remote medical training. Recently, telemedicine has been helped by advances in mobile communication and the adoption of tablets and smartphones. For example, it is currently easier to reach older or homecare patients whose medical access could be hampered by physical constraints. For example, mobile communications and portable devices permit monitoring of vital signs, no matter the patient's location.

Targeting one of the most critical chronic diseases, in 1985, there was an estimated 30 million adults living with diabetes mellitus. This number grew to 135 million in 1995, and then to 173 million in 2002 [1]. Currently, there are more that 371 million people living with diabetes between the ages of 20 to 79 years. This number rises every year in practically all countries, and 50% of the people with diabetes do not even know their clinical condition [2]. About two-thirds of these people live in developing countries, where the epidemic has more intensity, and with a growing rate in younger citizens [1]. Nevertheless, people with diabetes mellitus are living longer due mainly to advances in medicine [3].

As a consequence, both public and private healthcare costs for treating these types of patients are increasing due to higher average lifetime expectations worldwide. Additionally, both the number of medical doctors and their geographical distributions cannot cope with the quickly growing number of patients with chronic diseases. Finally, patients unaware of their clinical condition are not treated preventively and endure worse consequences since diabetes can be considered a “silent sickness”. They usually show greater resistance in following correct treatment since it is for the rest of their lives and demands daily care.

In Brazil, private health care operators offer Chronic Patients Management Programs to monitor patient clinical condition, mainly due to the possibility of income taxes reduction. In order to keep operational costs low, however, most of these programs only contact the patient once a month, either at home or through a phone call. If the patient does not feel well, he/she can call a monitoring center usually during working hours and speak with specialized nurses and doctors. This scenario limits the notion of the patient’s clinical condition. For patients with Type 2 diabetes, the medical protocol can lead to a possibly incorrect idea that they are clinically stable most of the time.

We present a mobile system – called Mobilicare – that is used by patients at home to collect and transmit patient data (e.g., glucose measurements) at specific dates and times. Through the system, the health center can remotely monitor a patient’s condition and even contact them. The main goals are to provide: a) a remote just-in-time pro-active and preventive health service to the patient, based on vital signs; b) graphs and alarms when vital signs are out of threshold for the health center and also the patient; c) video conference with doctors and family; d) patient risk analysis; and e) educational videos about diseases, food, and exercise habits. We observed that these features impact the way the treatment is performed, from reactive to active (i.e., anticipating an event). In order to analyze system outcomes over different scenarios, two case studies were conducted – one related to diabetes, and another to health promotion.

Methods

The Mobilicare System

The Mobilicare system is composed of three main entities, as depicted in Figure 1: a) the mobile kit at the patient's home; b) cloud storage; c) the medical/health care center.

Figure 1- General view of Mobilicare.

The mobile kit in the patient's home

The mobile kit is composed of a 3G/Wi-Fi tablet and a number of vital monitoring medical devices according to their chronic diseases (Figure 2). Electrocardiograms, scales, blood pressure meters, oximeters, thermometers, glucometers, pedometers, and spirometers are connected via Bluetooth to a tablet or smartphone. This device transmits information over the network to a Cloud storage system in which a medical center has access and analyzes such data for decision-making. The most common sensors are Glucometer (for diabetes), Blood pressure (for hypertension), and Scale (for weight control).

Figure 2. Medical devices supported by Mobilicare.

Figure 3 (left) shows a glucometer attached to a device that convert the measurement to Bluetooth, sending it to the tablet. For instance, the interface is configured only for diabetes measurements. Figure 3 (right) shows a graph of the vital signs – in this case for glucose – in which patients can follow and verify the disease stability over time.

Figure 3- Use for the tele-glucose meter (left) and glucose graphics (right).

Figure 4 depicts the video conference interface that includes the following features: access to the healthcare center, health coach for remote training, family, and a social network for people with the same disease. The goal of the social media part is so that patients can exchange with and motivate one another.

Figure 4- Mobilicare video conferencing and social network modules.

Cloud Storage

Mobilicare works as a service, i.e., it follows the SaaS (Software as a Service) technology and business model. The Cloud is used to store all the information and allow transparent web access through the Internet using RESTful APIs. Over the Cloud, the system both receives and automatically compares the measurements against a set of thresholds (upper as well as lower limits) prescribed by the responsible physician. These limits are individualized for each patient according to different clinical profiles.

By using cloud computing, medical professionals can communicate with several medical centers if needed via Web services. They can also manually analyze the vital signs (visualized using Java applets), perform a diagnosis, send alerts to the tablet application, and even contact patients via teleconference - all of that using a web browser at the medical center. The goal is not to replace a medical consultation, but rather complement and expand the service, offering guidance and monitoring from a distance.

The Medical Center

The medical center aggregates all patient data. If there is an alarm on a patient, their data appear at the beginning of the list and a sound is generated, followed by a yellow message. If the patient does not perform a prescribed measurement, the call center gets a "non-measurement alarm" and follows a protocol to question the patient for reasons for this behavior. If the measurement exceeds either a higher or a lower limit of the associated clinical profile, the call center both receives a "clinical alarm" (e.g., emergency, urgency) and executes the medical protocol.

Figure 5 shows the health care real-time dashboard as patients performance charts, contact, contact of close people, name of physicians, medicines took, and visit records, among others.

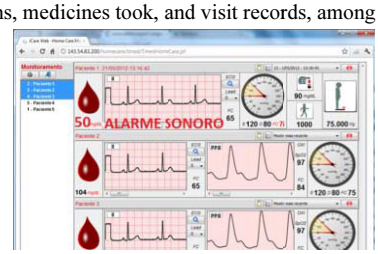


Figure 5- Mobilicare Medical Center dashboard.

Case study 1: 100 Diabetes participants over one year

Based on a Randomized Controlled Trial [7], 100 type two diabetes patients were classified into three distinct groups:

- GC (Control Group): received usual chronic disease management services only, which consisted of a call made by the health center attendant once a month to check patients' stability;

- GI1 (Intervention Group 1): used an offline glucose meter only, being managed by the patient himself;
- GI2 (Intervention Group 2): used a tablet with Android 4.0 and a glucose meter with a Bluetooth converter. These patients were monitored in real-time using Mobilicare by the healthcare provider.

Figure 6 shows the methodology workflow. First, 100 type two diabetes patients were selected from one of the largest home care providers in Brazil – Globalcare health center – and were chosen following the criteria of similar representation for all age groups and genders. Then, patients were randomized and distributed into the three subgroups (GC, GI1, and GI2). Eventually, a slight change was made among groups to keep a similar age distribution. Next, patients were asked to participate in the program (recruitment). In the case of a candidate refusing, another participant with a similar profile was chosen.

Once the set of patients were distributed into the three groups, a phase called "Intervention" took place. In this phase, patients had glycated hemoglobin collected for future comparisons. Afterward, the Follow-up, Data Collection, Data Processing, and Statistical Analysis phases were deployed (details in next section). The pilot lasted from October 2013 to September 2014.

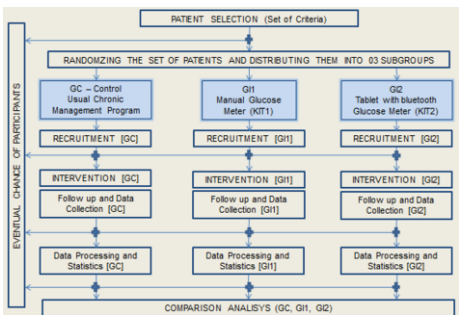


Figure 6- RCT – Randomized Control Trial Workflow.

Results

Diabetes

Figure 7 shows the monthly average levels of glucose measurements for the set of patients in GI2. It is possible to observe that the levels dropped from 118 to 111 using Mobilicare. For the other groups, there is no such a measure since they do not send their glucose levels to the health center. However, with the glycated hemoglobin levels reported, it was possible to verify the same behavior before the intervention.

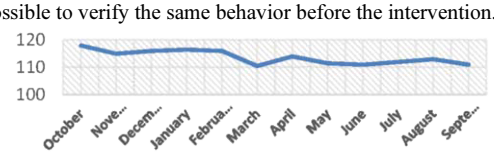


Figure 7- Mean value of glucose levels for GI2.

The number of "clinical alarms" among GI2 patients was around 62. Most alarms were triggered during typical vacation months, in which people often do not follow the specified diet. After a "clinical alarm", the average number of measurements to a GI2 patient to come back to an acceptable blood glucose level was 1.16, or a little more than two days. This quick return to a normal condition was a consequence of the medical protocols triggered by the alarms. For the other groups, they rarely reported an anomaly during the center calls.

Figure 8 shows the mean number of measurements for GI2. In the first months, there was instability due to system patient adaptability, which generated a high number of "non-measurement" alarms. Later, measurements became regular and stable with each patient performing about six measurements a week.

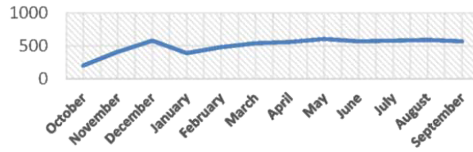


Figure 8- GI2 mean number of measurements per month.

The age average of the GI2 group is 69 years old, however this fact seems to have little influence in their adherence to the program, regardless of previous experience in using a tablet.

Also, one important result concerning the other two groups (GI1 and GC) is that less than 3% of the expected measurements were spontaneously reported to the "call center". This result can drive an assumption that these people could be under-watched and need a higher consciousness of the disease, as was provided for the patients of GI2.

Table 1 shows that the number of GI2 patients who went to the emergency room was 80% less when compared to GC patients in the same period. Also, GI1 patients were hospitalized 25% more times than GI2 patients and GC patients 50% more than GI2. The Unitcare Medical Services has details of the protocols, including ethical aprovement and agreement terms of patients.

Table 1: Emergency and hospitalization numbers.

	GI2	GI1	GC
Emergency Room	1	3	5
Hospitalization	4	5	6

Case Study 2: Health Promotion

This observational study intended to verify the Mobilicare performance for health promotion, i.e., the improvement of quality of life. This study has been carried out during a 4-week period, involving three people with different profiles:

- Patient 1: 56 years old, male, overweight and acquired diabetes type 2;
- Patient 2: 67 years old, male, overweight and Parkinson;
- Patient 3: 60 years old, female, overweight and hypertension.

During the first visit to the patient, a caregiver performed a training and also measurements for weight, blood pressure, and glucose. The results allowed the professional to get some indexes of the patients, like body mass index, body fat percentage, and waist to hip ratio. We also made an interview about their eating habits and physical activities during the last four weeks to better understand how the system could influence them.

After four weeks, the caregiver repeated the initial measurements and performed a final interview. After six months, another visit was made to verify self-health stability, without the Mobilicare influence. It is important to note that for statistical validation, this case study needs further investigation due to a restricted sample, but the preliminary results work as an indication of the system usability [15].

Health promotion

The patients in this study had to walk at least 5,000 steps daily and send the data, or else the system would show a "non-measurement" alarm to the health center. Figure 9 depicts the mean glucose and blood pressure variation of the patients. The glucose levels were high at the beginning of the experiments, and decreased to normal levels while the blood pressure levels, which were normal, remained normal. At the beginning of the experiment the glucose levels were about 120 in average, and at the end of the experiment they were about 100 in average, with a reduction of around 20%.

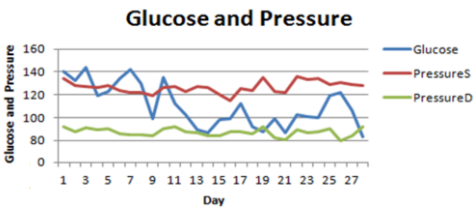


Figure 9- Patients mean glucose and pressure levels.

At the end of the experiment, the set of measurements was repeated. Table 2 presents the results for weight. All patients reduced their total weight, and additionally some of them increased in lean weight, reducing their percentage of body fat further. Fat weight and lean weight provide valuable information since they show that the intervention was probably producing significant results.

Table 2- Weight results.

	Total weight	Fat weight	Lean body weight	Body Fat (%)
P1_ini	85.00	16.14	68.86	18,99
P1_final	82.70	14.70	68.00	17.72
P2_ini	83.50	20.79	62.71	24.90
P2_final	82.00	16.50	65.50	20.12
P3_ini	73.10	25.56	47.54	34.96
P3_final	71.10	21.58	49.52	30.35

Other monitored results were the main corporal indexes, shown in Table 3: BMI (Body Mass Index), WHR (Waist to Hip Ratio) and BAI (Body Adiposity Index). The results showed a reduction the indexes for all patients. However, it is important to note that all participants still need to keep lowering most of their indexes in order to reach international health standards [4, 5, 6].

However, it is very hard to identify whether the benefits came from their motivation to adhere to the health program, and they internalized the learning, or by the pressure caused by the technology system itself. In order to minimize the error, the participants were contacted and measured again after six months of the experiment. Two of them had diminished the weight (6 kg and 2,3 kg), and the third was steady, with the diabetes controlled. The sample was only 3 participants, but the results are considered very promising.

Table 3- Main corporal indexes

	BMI	WHR	BAI
P1_ini	30.11	1.04	27.46
P1_final	29.30	1.01	26.54
P2_ini	29.94	1.00	29.72
P2_final	29.40	0.96	29.63
P3_ini	27.17	0.74	34.37
P3_final	26.43	0.74	33.51

Discussion

This paper presented Mobilicare – a new mobile monitoring for health promotion system. The front-end module works on any browser, being suitable for a variety of platforms, like computers, tablets, and mobile devices. One important goal was to encourage prevention and health promotion through the patient's interaction with advanced technological tools. In line with the concept of social networks, the developed tele-homecare model lets patients interact with other people who have similar medical situations, encouraging information exchange and commitment to self-care. Patients are encouraged to take an active role in their healthcare. For example, a diabetic patient could periodically check his blood sugar and adjust his or her behavior, such as in the case of an excessive intake of sugar. Additionally, the patient can optionally subscribe to an exclusive social network offered by the solution/health provider. There, patients can interact, post disease educational videos or materials for common access, discuss their health problems, and even compete with each other as they post their measurements and the disease remains stabilized over time.

Though it is still too early to be conclusive, we have collected significant evidence that remote measurement of vital signals, when integrated with awareness and healthcare support, might help leverage quality of life in the short term and delay the undesired side effects of chronic diseases like diabetes. Not only can the patient can get benefits using Mobilicare or similar systems, but also health care operators will probably be able to reduce emergency and therefore hospital costs.

We conducted two case studies, one for diabetes, and another focusing on health promotion. Related to the diabetes study, only 17% of GI2 patients have left the program after the end of the one-year experiment. Here, we observed that remote regular monitoring led to the earlier identification of more serious changes in glucose levels and also to an earlier reaction by health care personnel. As a consequence, patients' measurements were brought back to acceptable levels sooner. Related to the health promotion study, from the first day it was clear that motivation increased. Being constantly monitored brought a new mood to the participants. The system changed the routine of these people, and they worked out much more when compared to the time prior to monitoring. Now they had goals to reach, and someone to help them with their doubts and questions.

Results in both cases showed that a monitoring and educational system can improve the health of the monitored people. Moreover, we detected a real improvement in the knowledge of the patients about their health situation and what to do in order to grow old in a healthy way.

As counter-intuitive as it might be, elderly people can learn to use new technologies relatively quickly, and for a significant number of them the access to their measurement curves resulted in a positive feedback and led to greater adherence to the healthcare program. In this sense, the use of technologies that improve the quality of preventive health and reduce medical costs has become a necessity since the costs of secondary and tertiary care are very high and constantly growing. In BRICS countries, where there is significant heterogeneity in access to and quality of offered medical services, this market is expected to reach $418.4 million in 2014. In Brazil, it is estimated that less than 5% of hospitals are providing telemedicine services, indicating a high potential for expanded use of telemedicine solutions in healthcare.

Conclusion

Of course, data interoperability, security considerations, and privacy issues remain. Health information must be protected by the standard transmission protocols (Continua Health Alliance: http://www.continuaalliance.org) and security layers of the cloud provider to not only prohibit unauthorized access to patient records, but also to guarantee data integrity and make sure that transmitted data are not maliciously modified. Finally, privacy is addressed by the management system, which assures that data is only accessed by the physician and related medical team of the patient provided by the patient's permission.

References

[1] Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes. Estimates for the year 2000 and projections for 2030. Diabetes Care. 2004; 27(5):1047-53.

[2] 5th edition of the IDF Diabetes Atlas, 2012.

[3] World Health Organization. The World Health Organization Report 2002: reducing risks, promoting healthy life. Geneve, WHO, 2002.

[4] WHO. World Health Organization. 1995, 2000, 2004. BMI Classification. At: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Access: November, 2013.

[5] WHO. World Health Organization. 2011. Waist Circumference and Waist–Hip Ratio. At: http://whqlibdoc.who.int/publications/2011/9789241501491_eng.pdf. Access: November, 2013.

[6] Bergman, R. N., Stefanovski, D., Buchanan, T., Sumner, A. E., Reynolds, J. C., Sebring, N. G., Xiang, A. H., Watanabe, R. M. A better Index of Body Adiposity. In Obesity 19, 1083 – 1089, May 2011. At: http://www.nature.com/oby/journal/v19/n5/full/oby201138a.html. Access: November, 2012.

[7] Schulz, K. F., Altman, D. G., Moher, D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. BMJ: British Medical Journal. 2010. At: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844940/. Access: January, 2014.

[8] Eikerling, H-J., Grafe, G., Rohr, F., Schneider, W., 2009. Ambient Healthcare Systems - Using the Hydra Embedded Middleware for Implementing an Ambient Disease Management System. In Proceedings of the Second International Conference on Health Informatics. INSTICC Press, p. 82-89.

[9] Fensli, R., Boisen, E., 2008. Human Factors Affecting the Patient's Acceptance of Wireless Biomedical Sensors. In Communications in Computer and Information Science. Springer, v. 25, p. 402-412.

[10] Jaana, M., Paré, G., 2006. Home Telemonitoring of Patients with Diabetes: A Systematic Assesment of Observed Effects. In Journal of Evaluation in Clinical Practice. Blackwell Publishing, v.13(2), p.241-253.

[11] Pascual, J., Sanz-Bobi, M., Contreras, D. 2008. Intelligent System for Assisting Elderly People at Home. In Proceedings of the First International Conference on Health Informatics. INSTICC press.

[12] Sasaki, J., et al., 2009. Experiments of Life Monitoring Systems for Elderly People Living in Rural Areas. In Proceedings of the Second International Conference on Health Informatics. INSTICC Press.

[13] Sashima, A., et al., 2008. Toward Mobile Healthcare Services by Using Every Day Mobile Phones. In Proceedings of the First International Conference on Health Informatics. INSTICC Press, p. 242-245.

[14] Souidene W., et al., 2009. Multi-Modal Platform for In-Home Healthcare Monitoring (EMUTEM). In Proceedings of the Second International Conference on Health Informatics. INSTICC Press.

[15] Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J. and Stenild, S., 2006. It's Worth the Hassle! The Added Value of Evaluating the Usability of Mobile Systems in the Field. In Proceedings of the 4th Nordic Conference on Human-computer interaction: changing roles. ACM Press, p. 272 - 280.

[16] Kaufman, D.R., Patel, V.L., Hilliman, C., Morin, P.C., Pevzner, J., Weinstock, R.S., Goland, R., Shea, S. and Starren, J., 2003. Usability in the real world: assessing medical information technologies in patients' homes. Journal of Biomedical Informatics, vol. 36, Issues 1-2, pp. 45-60. Elsevier.

[17] Steventon et al. 2012. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ, 344:e3874.

Address for correspondence

Valter Roesler: roesler@inf.ufrgs.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-653

Calculation of Cardiac Kinetic Energy Index from PET images

John Sims[a], Marco Antônio Oliveira[b], José Claudio Meneghetti[b], Marco Antônio Gutierrez[a]

[a] *Division of Informatics, Heart Institute (InCor) Sao Paulo, Brazil*
[b] *Department of Nuclear Medicine and Molecular Imaging, Heart Institute (InCor) Sao Paulo, Brazil*

Abstract

Cardiac function can be assessed from displacement measurements in imaging modalities from nuclear medicine Using positron emission tomography (PET) image sequences with Rubidium-82, we propose and estimate the total Kinetic Energy Index (KEf) obtained from the velocity field, which was calculated using 3D optical flow(OF) methods applied over the temporal image sequence. However, it was found that the brightness of the image varied unexpectedly between frames, violating the constant brightness assumption of the OF method and causing large errors in estimating the velocity field. Therefore total brightness was equalized across image frames and the adjusted configuration tested with rest perfusion images acquired from individuals with normal (n=30) and low (n=33) cardiac function. For these images KEf was calculated as 0.5731±0.0899 and 0.3812±0.1146 for individuals with normal and low cardiac function respectively. The ability of KEf to properly classify patients into the two groups was tested with a ROC analysis, with area under the curve estimated as 0.906. To our knowledge this is the first time that KEf has been applied to PET images.

Keywords:

Kinetic Energy Index, cardiac function, optical flow, positron emission tomography, PET, Rubidium-82, Philips Gemini TF 64 TOF.

Introduction

Cardiac pathologies can affect the normal motion of the left ventricle; accurate quantification of this movement can therefore assist in diagnosing disease. The variational Optical Flow (OF) technique proposed by Horn and Schunck (HS) [1] can provide automatic motion quantification, in which the motion field is estimated from a sequence of images.

Optical flow is a computer vision technique which estimates the motion field from spatiotemporal patterns of image intensity [2], and HS requires that the image brightness, E, at a given point in the brightness pattern does not change with time, t, as shown in Equation 1.

$$\frac{dE}{dt} = 0 \tag{1}$$

The motion field is estimated through the solution of a system of equations which is large and sparse, and iterative methods such as Gauss-Seidel are suitable choices for its solution [1]. In 4d images (a time sequence of 3d volumes) these equations have the form shown in Equation 2; for brevity only the x component of velocity (u) has been shown.

$$u^{k+1} = \bar{u}^k - \frac{E_x(E_x\bar{u}^k + E_y\bar{v}^k + E_z\bar{w}^k + E_t)}{\alpha^2 + E_x^2 + E_y^2 + E_z^2} \tag{2}$$

Here k is the iteration index; $\bar{u}, \bar{v}, \bar{w}$ are mean velocities in x, y, z; E_x, E_y, E_z are the partial derivatives of image intensity E(x,y,z,t) and α^2 is a regularization parameter.

The motion field estimated from cardiac image sequences is difficult and time consuming to interpret, since each volume element (voxel) has a time-varying three-dimensional vector value. Gutierrez et al. [3], applied optical flow techniques to estimate the velocity field from single photon emission computed tomography (SPECT) images and then used the velocity estimates to calculate kinetic energy (KE) for each voxel of the image using Equation 3.

$$KE = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) \tag{3}$$

Here v_x^2, v_y^2, v_z^2 are the components of the velocity vector in x, y and z directions respectively for a voxel at position (x,y,z), and m is the voxel mass, approximated by the voxel intensity in the 3D volume image [4].

The resulting KE field can be summed within each frame to provide total KE per frame and then represented graphically over the cardiac cycle to produce a KE curve. An indicator of cardiac function was proposed in [3], called kinetic energy index (KEf), shown in Equation 4.

$$KEf = \frac{KE_{max} - KE_{min}}{KE_{max}} \tag{4}$$

where KE_{max} and KE_{min} are maximum and minimum values of the total kinetic energy per frame respectively [3].

The present study has the objective of evaluating KEf for the classification of patients with normal or low/altered cardiac function from PET images.

The images come from an image database of 666 individuals with normal and low/altered cardiac function who underwent a rest and stress examination in the Department of Nuclear Medicine at the Heart Institute, Sao Paulo. Images were acquired with the PET radioisotope Rubidium-82 (^{82}Rb).

Methods

Image acquisition

After being situated in the Philips Gemini PET/CT scanner[1], ^{82}Rb was administered for 60 seconds. The individual was left for 60 seconds for perfusion of ^{82}Rb, then resting acquisition performed from minutes two until eight. Cardiac stress was then induced pharmacologically, ^{82}Rb administered a second time during 60 seconds, followed by another 60 second wait for perfusion, and finally the stress image acquired from minutes two until eight.

The electrocardiogram (ECG) signal was simultaneously recorded by the Philips Gemini during image acquisition (gated PET) so that a sequence of image volumes could be constructed from detected photon events collected over several hundred RR intervals. Each RR interval defines the times between R peaks of the ECG, and relates to a single

[1] Philips Gemini TF 64 TOF, Series number 7535, Manufactured Jan 2012 by Philips Medical Systems (Cleveland) Inc., Cleveland, Ohio, USA, Identification code MN 0183, HC 147938

complete cardiac cycle; the length of RR intervals vary constantly during the examination as the cardiac frequency of the individual varies.

Sixty three images were selected from the image database to create groups of individuals with (i) normal cardiac function (n=30) and (ii) low/altered function (n=33). For this preliminary study, the criterion for inclusion in the normal group was ejection fraction (EF) > 50 and the low/altered group was EF < 50, both as determined by the stress examination. Figures 1 and 2 show the variation in EF for individuals as measured by stress and rest examinations.

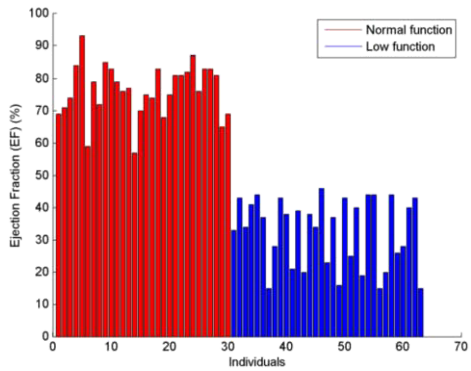


Figure 1– Ejection fraction in stress for the 63 individuals

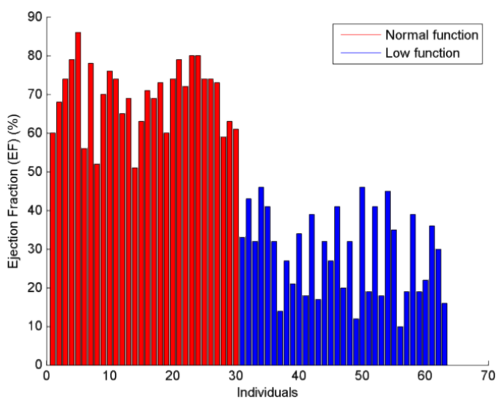


Figure 2– Ejection fraction in rest for the 63 individuals

Calculation of KEf

We developed software in MATLAB[2] to produce an automated image processing routine for calculating KE*f*, comprising stages of preprocessing, optical flow calculation, and postprocessing as shown in Figure 3.

Input images were a sequence of 8 volumes, with each volume containing 128 slices, each of which is 128x128 pixels. The slice thickness was 4mm and pixel size 4mm x 4mm. 8 gated frames were produced rather than 16 since the acquired count is low for ^{82}Rb and it was thought that spreading the count over 16 frames would cause unacceptably high levels of noise.

[2] MATLAB Release 2013a, The MathWorks Inc., Natick, Massachusetts, United States

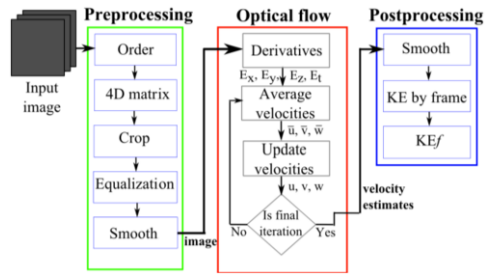


Figure 3– Scheme for calculation of KEf

The preprocessing stage comprised the following:

- Order: DICOM images were renamed in sequential order from the first slice of the first volume to the last slice of the last frame.
- 4D Matrix: DICOM images were converted and stored in a binary MATLAB 4D matrix format.
- Crop: An automatic crop was performed to reduce the size of the input image volumes to include only the left ventricle (LV). A Hough circle transform was used to locate the LV in slices of the short axis plane, cropping the image to include a minimum of structures not belonging to the LV
- Equalization: Images were adjusted to equalize total brightness per frame after cropping, as detailed below.
- Smooth: Images were low pass filtered by averaging over the 27 neighbourhood of each voxel.

The optical flow stage of processing used a version of the HS scheme adapted to 4d images, as described in [2]. A total of 200 iterations of the algorithm were performed with a value of α^2=10, found by experiment.

- Derivatives: Partial derivatives of image intensity were calculated using centred differences.
- Average velocities: In each iteration of the algorithm, average velocities were calculated for each component of velocity (u,v,w), for each voxel, over its 18 neighbourhood.
- Update velocities: an update step was applied to produce the new values of velocity component (u,v,w).

The post-processing stages were:

- Smooth: The optical flow field was median filtered in the 27 neighbourhood to remove outliers.
- KE by frame: KE was calculated for each voxel using Equation 3 and then KE contributions summed for all voxels within each frame.
- KEf: KEf was calculated using Equation 4 from maximum and minimum values of KE by frame. Our studies excluded the values belonging to the first and last frame from this calculation since these frames have a single neighbour in time, increasing the error in the estimate of partial derivative in time.

After an analysis of initial results, it was discovered that the total brightness (i.e. photon count) of the image varied substantially between frames. This was unexpected since the number of photons emitted at any given moment is dependent only on the random radioactive process. In order to quantify brightness variation, the vector B_{total} was formed whose elements represent the total brightness in each frame of the sequence, as shown in Equation 5.

$$B_{total} = [\sum_{v=1}^{X} E_v^1, \sum_{v=1}^{X} E_v^2, \ldots, \sum_{v=1}^{X} E_v^N \ ] \quad (5)$$

where N is the number of frames in the sequence, X is the number of voxels in each frame, E is the intensity of the vth voxel. Let f be the frame index, so that the fth element of B_{total} is B_{total}^f.

Total brightness per frame was analyzed over the cardiac cycle and a brightness equalization module added to make total brightness in each frame constant over the cardiac cycle, as shown in the preprocessing phase of Figure 3. The algorithm followed the steps:

1. Form the vector B_{total} for all frames .
2. Identify the maximum element of B_{total}, max(B_{total}).
3. Calculate new intensities using Equation 6.

$$E_{NEW(i,j,k)}^f = E_{(i,j,k)}^f \cdot \frac{\max(B_{total})}{B_{total}^f} \quad (6)$$

KE curves, showing the variation of total kinetic energy by frame, were plotted and KEf calculated for individuals with normal (n=30) and low/altered cardiac function (n=33) in rest and perfusion images. KE Curves were produced with and without the brightness equalization step so that any improvements created by the new module could be assessed.

Statistical analysis

In order to investigate the relationship between EF and KEf in rest, linear regression was performed on the pairs of values (KEf, EF) formed for each individual.

In order to evaluate the ability of KEf as a classifier of cardiac function, sensitivity and specificity tests were performed. Sensitivity was defined as the fraction of individuals with normal function who were correctly classified by KEf and specificity as the fraction of individuals with low/altered function who were correctly classified by KEf. A receiver operating characteristic (ROC) curve [5] was generated by varying a KEf threshold, T, from 0.01 to 1.0 in intervals of 0.01. For each threshold, the number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) were calculated, and from these values true positive rate (TPR), false positive rate (FPR) and Accuracy (A) calculated as described in reference [5]. Accuracy is the total number of true classifications as a fraction of the total number of patients. The area under the ROC curve (AUC) was estimated using the trapezoidal rule.

Results

Values of KEf were generated for 63 individuals with normal and low/altered cardiac function.

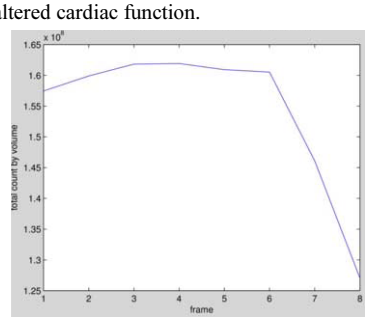


Figure 4– Brightness curve for an individual with low cardiac function (Rest EF = 33%)

Figures 4, 5 and 6 were produced for one of the 33 individuals with low/altered cardiac function and ejection fraction at rest of 33%.

Figure 4 shows the total brightness curve, which has a sharp brightness decrease in both frames 7 and 8. Figures 5 and 6 show the KE curve without and with brightness equalization performed in the preprocessing stage. The large peak in frame 7 of Figure 5 is greatly reduced in Figure 6.

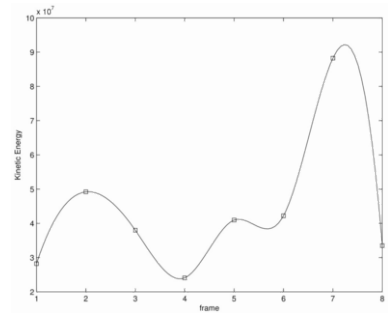


Figure 5– KE curve without brightness equalization (KEf = 85.69%)

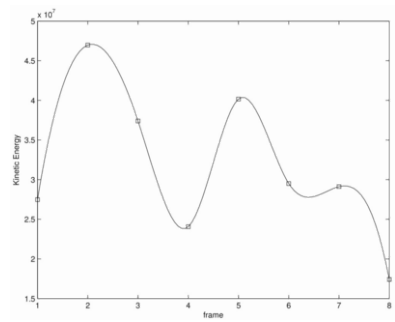


Figure 6– KE curve with brightness equalization (KEf = 50.15%)

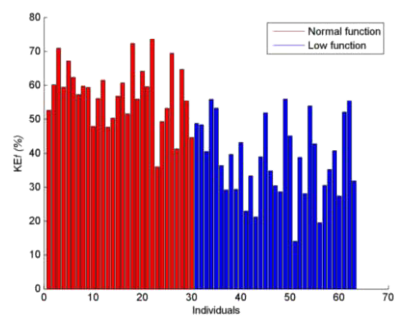


Figure 7– Bar chart showing KEf for rest perfusion images for the 63 individuals in the study

Figure 7 shows the values of KEf calculated for every individual in the study. It can be observed that individuals with normal function have higher KEf than those with low or altered function in general.

Mean and standard deviation of KEf calculated for these two groups and are shown in Table 1. The mean values of the two groups are separated by 19.2%.

Table 1– KEf values obtained from rest perfusion images

	KE*f* in Rest
Normal (n=30)	0.5731±0.0899
Low/altered (n=33)	0.3812±0.1146

Table 2 is a frequency table showing the frame in which minimum brightness occurred for individuals with normal and low/altered cardiac function. The most common frame was frame 8, with a total of 46.0% over the two groups.

Table 2– Frequency table showing the frame in which minimum brightness occurred for the two groups of cardiac function

Frame	1	2	3	4	5	6	7	8
Frequency (normal)	4	2	2	3	3	2	3	11
Frequency (low/altered)	2	3	1	2	2	2	3	18

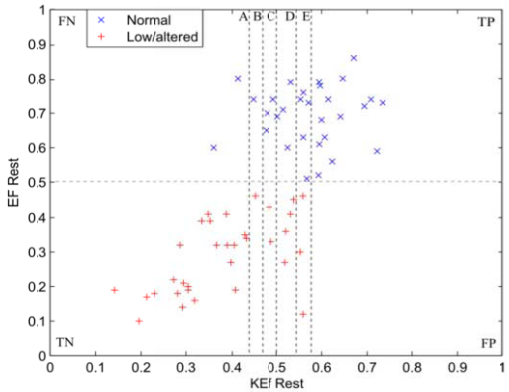


Figure 8– Plot of EF against KEf for the 63 individuals

Statistical analysis was performed on the data. Figure 8 shows a plot of the values obtained for KEf at rest against EF at rest. Regression analysis was performed using MATLAB Statistics Toolbox v.8.2, finding an intercept with the vertical axis of -0.077 and gradient of 1.185. R-squared and adjusted R-squared were 0.546 and 0.538 respectively, with a P-value of 4.79e-12.

Figure 8 is divided into two clusters, with the blue 'x' markers showing the group with normal cardiac function and the red '+' markers showing the group with low or altered function. A horizontal line at EF at rest of 50% divides the two groups. Five thresholds, marked A,B,C,D,E, were manually drawn, and TPR and FPR calculated for each in order to test the ROC curve generated in MATLAB. A summary of the test results is shown in Table 3, where the threshold letter is shown together with the threshold value in the first column. The quadrants of Figure 8 have been labeled as TP, FN, FP, TN, (respectively top-right, top-left, bottom right, bottom left) to indicate how testing was performed; the number of detections in each of the four quadrants were summed for each threshold A, B, C, D, E.

Table 3– Test of thresholds used in ROC curve

T	TP	FN	FP	TN	TPR	FPR	A
A (0.43)	28	2	23	10	0.933	0.333	0.81
B (0.46)	27	3	9	24	0.900	0.273	0.81
C (0.50)	24	6	7	26	0.800	0.212	0.79
D (0.54)	20	10	3	30	0.667	0.091	0.79
E (0.58)	15	15	0	33	0.500	0.000	0.76

Figure 9 shows the ROC curve for KEf. AUC was determined to be 0.906 and placing the threshold at KEf = 49% creates the point in the ROC curve closest to the point (0,1) with a distance of 0.270. Thresholds A-E have been marked and a manually fitted curve drawn.

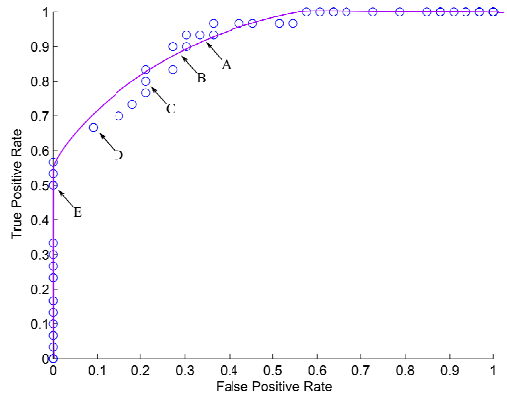


Figure 9–ROC curve for sensitivity and specificity test for classifying normal cardiac function based on KEf

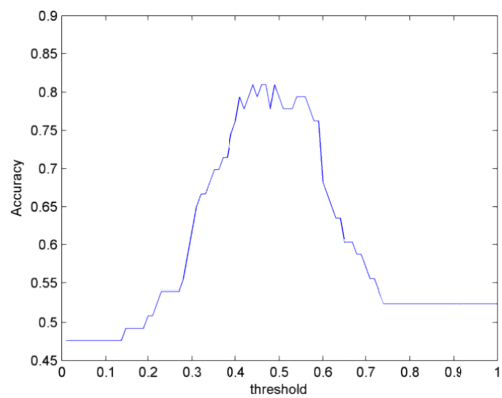


Figure 10– Accuracy against threshold

Figure 10 shows the value of accuracy as the threshold is varied. Accuracy has a maximum value of 0.810 when the threshold is in the region 0.43 to 0.49.

Discussion

Gutierrez et al. [3] showed the expected form for KE curves by averaging KE per frame over 30 individuals with normal cardiac function in SPECT studies. The KE curve seen in Figure 6 has a similar form, with the global maximum in frame 2, a local maximum in frame 5 and a local minimum of velocity between these peaks in frame 4. The temporal resolution is not sufficient to precisely identify cardiac phases. The exact frames in which these events are observed was seen to vary between individuals.

The KE curve without brightness equalization, Figure 5, shows a very large peak in the seventh frame which is not seen after equalization, Figure 6. The corresponding brightness curve, Figure 4, has a sharp drop in brightness in the eighth frame, which corresponds to the position of the large peak. As we showed in Equation 1, the HS method requires objects in the image to have constant brightness, and errors in velocity estimation occur when brightness of objects varies. Since brightness equalization causes the large peak of the KE curve to disappear, giving rise to a curve more similar to those shown in [3], we conclude that the brightness change causes the large peak in Figure 5 and that the equalization module is

able to improve the suitability of the image sequence for the calculation of KEf in this case.

Extra or altered maxima and minima in the KE curve can cause errors in the value of KEf calculated. The calculated value of KEf from Figure 5 and 6 were very different, with values of 85.69% and 50.15%; the global maximum without brightness equalization, from Figure 5, belongs to the large peak, but with brightness equalization, Figure 6, to the first smaller peak. This decrease in KE_{max} reduces the value of KEf.

The cause of the drop in brightness is not easy to ascertain since the formation of the image takes place within proprietary hardware and software. We believe the origin of this phenomenon lies in the process of image construction from varying lengths of RR intervals, since those individuals with the largest variation in cardiac frequency during acquisition also tend to have the greatest variation in brightness by frame and the minimum brightness is most likely to occur at the end of the sequence (Table 2). We hope to better understand how to avoid this change in brightness so that our images are more suitable for motion estimation.

Table 1 shows KEf mean and standard deviation for resting perfusion images of individuals with normal and low/altered cardiac function. There is a separation of 19.2% between the means of the two groups. Figure 7 shows KEf results in bar graph form, showing some overlap between the two groups in terms of KEf.

The selection of individuals for this preliminary study was performed using a simple criterion which could produce mixed groups in terms of cardiac function, reducing separation between the groups and increasing standard deviation. After discussion with medical experts, more rigorous criteria were suggested to categorize the individuals in the image bank, combining measurements of ejection fraction, perfusion, and coronary blood flow. It was also suggested that the image bank be divided into subcategories, perhaps based on a particular cardiomyopathy or degree of disease, as assessed by experts in nuclear medicine.

The HS method requires small voxel displacements between frames and a temporal resolution of 8 frames per cardiac cycle means that the voxel displacements could be inappropriately large. Reconstruction could be performed with 16 frames, but the low count in the Rubidium images means that the signal to noise ratio might become unacceptably low, since the HS method is also sensitive to noise.

The brightness equalization was seen as a necessary step to eliminate inaccuracies in the optical flow estimation; however noise in frames which have low total brightness will be amplified and could cause errors in the estimation of partial derivatives and subsequently in the velocity estimate.

Linear regression shows a gradient close to 1.185 with an intercept on the vertical axis close to -0.077. The R-squared values show that there is some correlation EF and KEf, but it is not very strong. There is a definite trend of KEf increasing as EF increases, but the dispersion of the points is quite large and there are several outliers.

AUC provides the probability that a classifier will rank a positive instance, chosen at random, higher than a negative instance, chosen at random. An AUC of 0.5 means the classifier result is only as good as a guess [5]. Here the AUC is 0.906, indicating that KEf rest has performed reasonably well in classifying individuals with normal from low/altered cardiac function. It should be noted that the value of AUC for KEf at rest does not relate directly with correlation with EF at rest.

Conclusion

In this study KE index (KEf) was evaluated as an indicator for classifying individuals in normal and low/altered cardiac function groups from PET images, with the study showing reasonable results for its ability.

Variation in brightness through the image sequence was seen to cause inaccuracies in the estimation of velocity, which in turn causes significant errors in KEf. We showed how equalization of brightness can reduce errors in KEf.

Future work will involve a more rigorous selection of individuals in the normal and low/altered cardiac function groups, including larger group sizes, and comparing KEf results obtained from images formed from 8 and 16 gated frames.

The results presented were obtained from perfusion rest image sequences, but future investigations could include calculating KEf for the stress image sequences.

The ability of KEf to properly classify patients into the cardiac function groups was tested with a ROC analysis, with area under the curve estimated as 0.906.

To our knowledge this is the first time that kinetic energy index has been applied to PET images.

References

[1] Horn BKP Schunck BG. Determining optical flow, Artificial intelligence, 17:185-203, 1981

[2] Barron JL, Thacker NA. Tutorial: Computing 2D and 3D Optical Flow, Tina Memo No. 2004-012, 2005

[3] Gutierrez MA, Rebelo MS, Furuie SS, Meneghetti JC. Automatic quantification of three-dimensional kinetic energy in gated myocardial perfusion single-photon-emission computerized tomography improved by a multiresolution technique. Journal of Electronic Imaging vol. 12(1) 118-124, 2003

[4] Gutierrez MA, Rebelo MS, Furuie SS,Pozzo L, Meneghetti JC,Melo CP, The Assessment of Myocardial Motion in Gated-SPECT using a Multiresolution Technique, Computers in Cardiology, 27; 303-306, 2000

[5] Fawcett T, An introduction to ROC analysis, Pattern Recognition Letters, 27;861-874, 2006

Address for correspondence

Mr John Sims

Division of Informatics, Instituto do Coração, HCFMUSP,

Av. Dr. Enéas de Carvalho Aguiar, 44

CEP: 05403-900 - São Paulo/SP,

Brazil

Email: john.sims.incor@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-658

TCGA4U: A Web-Based Genomic Analysis Platform To Explore And Mine TCGA Genomic Data For Translational Research

Zhenzhen Huang[a], Huilong Duan[a], Haomin Li[b,c]

[a]*College of Biomedical Engineering and Instrument Science, Zhejiang University, China*
[b]*The Children's Hospital, Zhejiang University, China*
[c]*The Institute of Translational Medicine, Zhejiang University, China*

Abstract

Large-scale human cancer genomics projects, such as TCGA, generated large genomics data for further study. Exploring and mining these data to obtain meaningful analysis results can help researchers find potential genomics alterations that intervene the development and metastasis of tumors. We developed a web-based gene analysis platform, named TCGA4U, which used statistics methods and models to help translational investigators explore, mine and visualize human cancer genomic characteristic information from the TCGA datasets. Furthermore, through Gene Ontology (GO) annotation and clinical data integration, the genomic data were transformed into biological process, molecular function, cellular component and survival curves to help researchers identify potential driver genes. Clinical researchers without expertise in data analysis will benefit from such a user-friendly genomic analysis platform.

Keywords:

Data Mining; Genomic Alterations; Driver Genes.

Introduction

Many large-scale human cancer genomics projects offer huge quantities of genomic data from tumor samples, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). TCGA provides a platform for researchers to search, download, and analyze datasets including clinical information, genomic characterization data and high level sequence analysis of the genomes of nearly 50 tumor types [1]. Many researchers use statistical methods and algorithms on clinical and high throughput genomic data, including copy number alterations (CNAs), mRNA and small RNA expression, somatic mutation, and methylation data to find potential driver mutations and genes to improve the prevention, early detection, and treatment of cancer [2-9]. There are many clinical researchers without enough data analysis and bioinformatics training who face an embarrassing situation if they don't have the professional abilities to appropriately manipulate the downloaded data. In recent years, web-based analysis tools such as the Cancer Genome Workbench, Integrative Genomics Viewer, cBioPortal for Cancer Genomics, and Broad Firehose have been used by clinicians and researchers to search meaningful genomic alterations to develop targeted and personalized clinical treatments [1, 9].

To help clinical cancer researchers fully benefit from the TCGA datasets through a simple and user-friendly tool, we developed a web-based platform named as TCGA4U. TCGA4U is an intuitive web-based analysis tool to analyze high-level genomic data of different TCGA samples in distinct cancer types. Currently, the platform offers statistical analysis results and graphical views to help users find interesting results for further investigation. Besides providing the specific gene or gene list, it also provides a genomic characteristics query service, such as CNAs, Somatic Mutation, Gene Expression and DNA Methylation. Furthermore, TCGA4U also integrates the clinical data, the Gene Ontology and data mining results with the gene level data to provide more insights for clinical investigation.

Materials and Methods

TCGA4U Database

Genomic Data

All the genomic analysis data in TCGA4U were downloaded from TCGA's open-access data tier. In this study, two cancer types (Lung adenocarcinoma (LUAD) and Lung squamous cell carcinoma (LUSC)) were used. For each cancer type, TCGA has multiple data types and data levels available for open access, such as Copy Number, Somatic Mutation, and DNA Methylation. Five types of data were used in this study:

1. Copy Number Variants (CNVs) datasets which account for at least 10% of genetic variation and CNVs amplification and deletion may be responsible for many cancers [10, 11]. The level 3 CNVs datasets of LUAD and LUSC were downloaded on Feb 18, 2014.
2. Somatic mutations datasets provide important genomic information and are essential for tumor occurrence [3]. The somatic mutation datasets of LUAD and LUSC were downloaded on Apr 16, 2014.
3. Expression-gene datasets provide interesting expression patterns under different biological conditions for investigators. The expression-gene datasets of LUAD were downloaded on Mar 31, 2014. The LUSC datasets were downloaded on Apr 3, 2014.
4. DNA methylation datasets play a critical role in the regulation of gene expression. LUAD DNA methylation datasets were downloaded on Feb 18. LUSC datasets were downloaded on Apr 4, 2014.
5. Clinical data which contain follow-up information were downloaded on March 19, 2014.
6. All these downloaded datasets were imported into relational database tables defined by expanding the columns defined in the downloaded tab-delimited files.

Genome Reference Data and Other Knowledge Data

The February 2009 human reference sequence (GRCh37) was downloaded from the UCSC Genome Browser. These data provided location information of genes to explore the genomic data such as CNVs and somatic mutations.

In the Gene Ontology (GO) term mapping module, we used the GO online database of the European Bioinformatics

Institute (EBI) to generate directly utilizable gene mapping files, term nodes and edge files. Specific GO term background rates in the genome were calculated and stored as local files. The relationships between GO terms were stored using term nodes and edge files to generate networks.

The EBI Molecular Interaction Database (IntAct) was downloaded on Jun 5, 2014. These data provided a related gene list of specific genes that will be used to identify molecular interaction relationships.

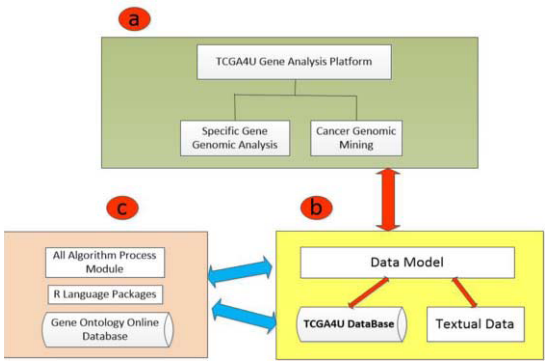


Figure 1 – TCGA4U System Schema

TCGA4U Platform

As shown in Figure 1, the TCGA4U platform provides two functions for the user. The first one is to provide a direct genomic query service that allows the user to query genomic characteristics of specific genes under different cancer types in the TCGA4U datasets. The second is a data mining result that allows the user to explore the driver genes for targeted cancer types in the GWAS level.

Genomic Analysis of Specific Gene

Four types of genomic characteristics with clinical data were available for genomic analysis. Clinical researchers only need to select the interested cancer type and input the gene symbol to obtain analysis results with four parts: CNVs, DNA Methylation, Somatic Mutations, Gene Expression.

Copy Number Variants Analysis

The CNVs datasets of tumor and matched normal were stored separately in two database tables. The gene location information obtained from the genome reference data will be used to obtain the CNVs (log_2-fold change) of this gene from these tables. A histogram of CNVs' values (as shown in Figure 2.A) will give the user an overview of this gene's CNVs under this cancer. Additionally, CNVs' mean, variance, and *p*-value (t-test) of tumor and normal samples can be calculated and shown to the user. Overall, these data can illustrate the amplification or deletion status of the gene copy number.

Integrating with clinical follow-up datasets, patient groups that are separated by CNVs results (level of amplication or deletion) will give four survival curves by Kaplan-Meier estimator (as shown in Figure 2.B).

DNA Methylation Analysis

The beta values of DNA methylation at each CpG site will be plotted as error bars as shown in Figure 2.C. More statistics including minimum, maximum and deviation will be also calculated to assist the user in assessing the DNA methylation rates in specific sites.

Furthermore, the DNA methylation rates at specific CpG sites can also separate patients into two groups (high or low beta value) which can be used to plot different survival curves for the user.

Somatic Mutations Analysis

The number of mutation samples of 11 somatic mutation types at different sites was plotted as a scatter diagram with 11 different colors and shapes (as shown in Figure 2.D). The user can easily identify the hotspot that has a high mutation rate in this mutation type of cancer. Some basic statistics, such as the number of mutation samples, total number of samples, and mutation rate were also calculated to reflect the mutation conditions of the gene in this cancer.

As we have done for other data types, mutation status of specific genes or sites were used to separate patients into different groups to show different survival curves to explore the influences of somatic mutations of specific genes on patients survival.

Gene Expression Analysis

Gene expression data (log_2-fold change) of specific genes were displayed in a histogram as shown in Figure 2.E. Using the IntAct data, a gene list related with query genes will be generated. An expression heat map of this gene list in different samples will be generated by using the R programming language (as shown in Figure 2.F). To make it more flexible, users are allowed to customize the gene list to generate a corresponding heat map. The visualized data and clusters will better illustrate the potential interactions in these genes.

As with other data types, levels of expression divide patients and generate survival curves, revealing the influences of different levels of expression on patients' survival.

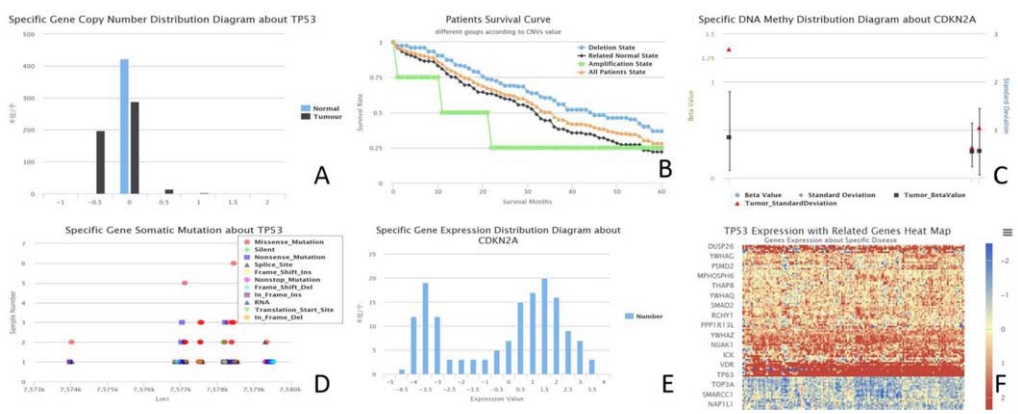


Figure 2 – Specific gene genomic analsysis. A. CNV analysis; B. Survival Curves of patients seperated by CNVs; C. DNA Methylation analysis; D. Somantic Mutation Analysis; E. Gene Expression Analysis; F. Heatmap of a gene list.

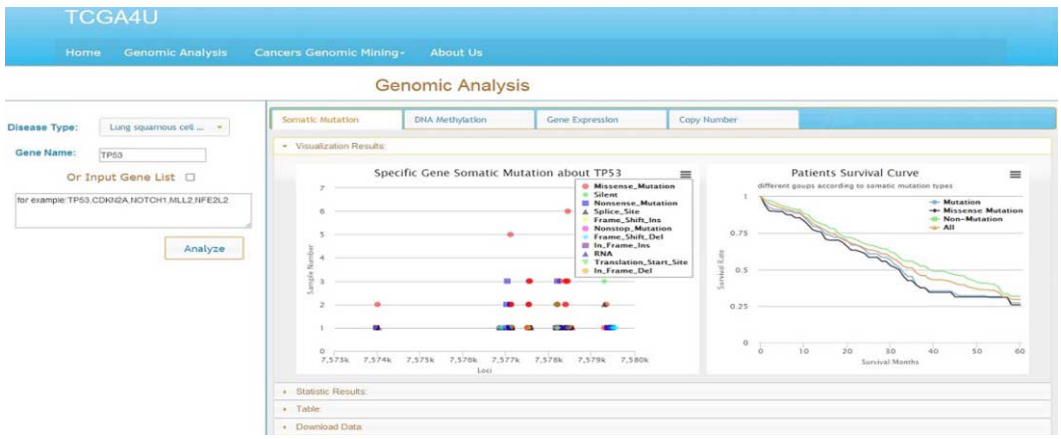


Figure 3 – Specific gene genomic data analysis page

Potential Driver Genes and Mutations Mining

The ability to identify driver genes or prioritize genes from TCGA datasets will help clinical investigators find these target genes for further research. An approach proposed by Trifonov named MutComFocal was used in a TCGA platform [12]. In brief, the MutComFocal algorithm is a Bayesian approach to identify candidate driver genes by integrating point mutation and copy number information in a large number of tumor samples. This method takes into account the recurrence of altered genes with the size of the lesions. Each gene will have calculated three basic and four composite score types: focality score, recurrence score, mutation score, amplification score, deletion score, amplification-mutation score (amp&mut), and deletion-mutation score (del&mut). Genes have separately calculated scores for copy number deletions and gains, with the assumption that high amplification scores will tend to be oncogenes and, conversely, high deletion scores will tend to be tumor suppressors. Finally, a mutation score separately combines with the amplification and deletion score to accomplish somatic mutations combining with CNVs. After that, genes can be tiered based on score distribution. In general, top tier genes are more concerning to us, and more likely to be driver genes. In TCGA4U, a mining tool based on this algorithm was developed. The LUAD and LUSC data was analyzed. A genome level visualization tool was developed to show the gene tiers, score and locations.

Gene Ontology Annotation analysis

The high priority genes in the first tier always contain more than several hundred genes. To further narrow-down the list, a GO annotation enrichment analysis was provided on the platform.

In this process, each gene in the list will be annotated with GO terms in three domains: Cellular Component, Molecular Function, and Biological Process [13]. The terms usage in the list will be compared with the genome background usage rate (the proportion of the number of gene products which mapped to a specific term on total number of gene products) to highlight some biological process that are involved in tumor development. Cytoscape.js – a large-scale network javascript library for graph analysis and visualization – was used to show the GO term networks. The interested terms will track back to the top level terms in the visualized network diagram with enrichment annotation to help the user understand the terms relationships [14].

Results

A publicly accessible website was developed based on these methods and published at http://www.tcga4u.org:8888.

Specific Gene Analysis: genomic alterations characteristics analysis

As previously mentioned, we developed a Specific Gene Genomic analysis module in TCGA4U platform as shown in Figure 3. The left panel was the user operating area that allows the user to choose specific cancer types and input a gene symbol as a query condition. The right part is the main display area. By switching between tab pages, different genomic data type results will be displayed for the user. Each tab page contains some accordion display modules (such as Visualization Results, Statistic Results, Original Table, or Download Data) to help the user locate interesting analysis results. In general, Visualization Results offered corresponding genomic type visualization results and survival curves. Statistic Results mainly provided some basic statistic results such as mean, variance, and the *p*-value of CNVs. Original Table and Download Data were for the convenience of the user to query and download raw data. Additionally, the Gene expression page also offered a heat map of related gene lists. Overall, this function module was designed to offer a simple and user-friendly operation and interface to help the user obtain meaningful analysis results about genomic alterations.

Offer Potential Driver Genes and Tiers Display

To offer researchers some valuable research objects, we developed the Potential Driver Genes Mining module. The same as the Genomic Analysis of Specific Gene module layout, the operating area made the user choose a specific disease and algorithm as query conditions. The right part contained three tab pages: Gene Tier Charts, Gene Score Tables, and Diverse Diseases Complex Heatmap. Gene Tiers Charts displayed Amplified and Deleted Gene Tiers and Amplified&Mutation and Deleted&Mutations Gene Tiers (as shown in Figure 4). The horizontal axis represented score value and vertical axis represented chromosomal location information. Genes were plotted as points with different colors meaning different tiers for the convenience of viewing the distribution of different tier genes on the chromosomes. Meanwhile, each chart marked the top 20 genes to facilitate the user viewing.

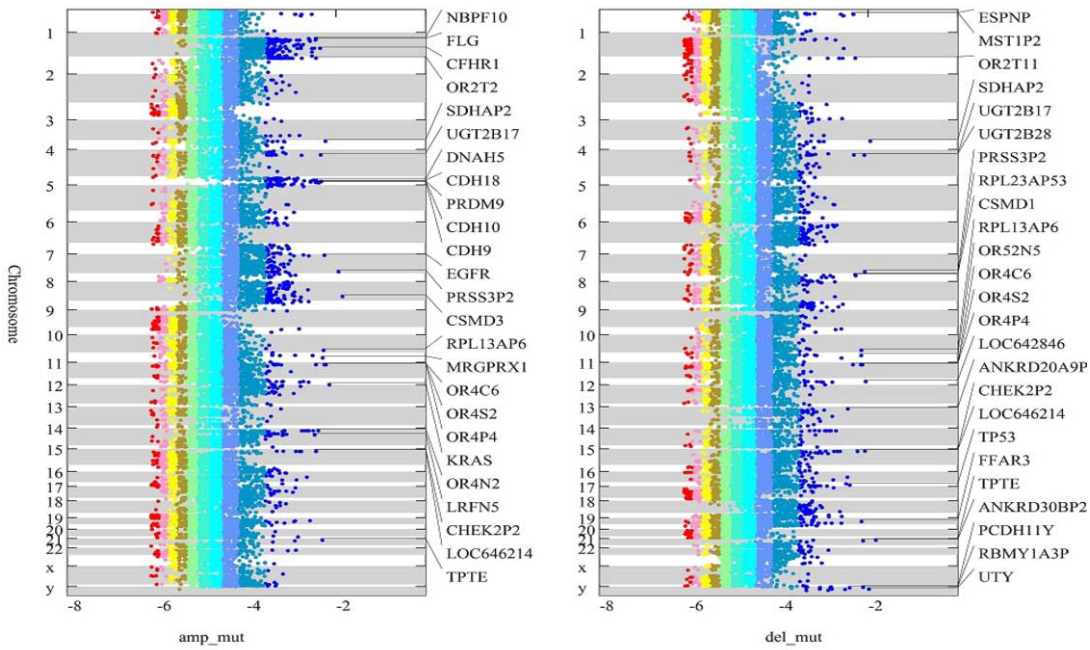


Figure 4– Amplified&Mutations and Deleted&Mutations Gene Tiers of LUAD

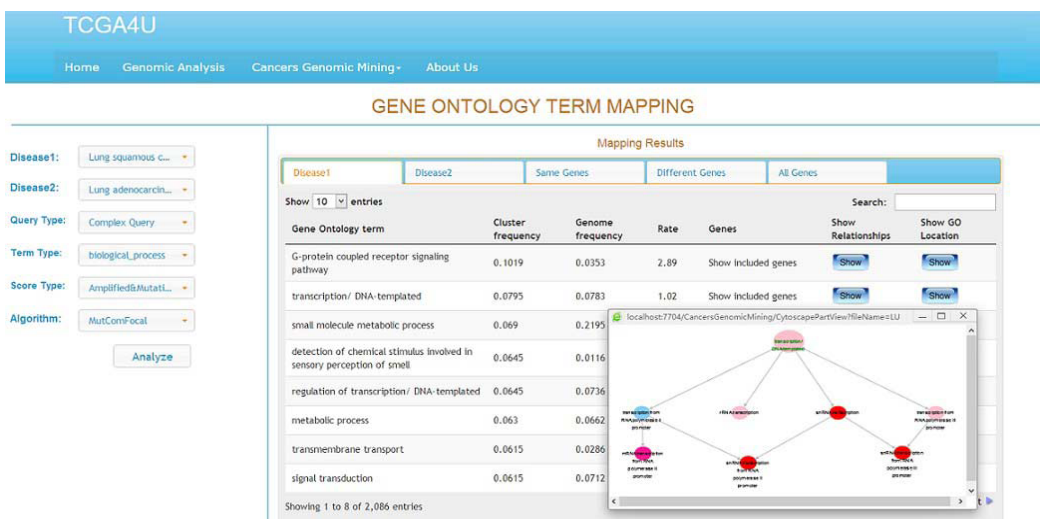


Figure 5– Gene Ontology Term Mapping Analysis Page

GO Term Mapping Analysis

A GO term mapping analysis page was developed as shown in Figure 5. In the query area, the user can choose one or two cancer types, algorithm type, query type, score type, and term type. In the diplay area, there is a GO term mapping table offering specific term mapping frequency, background mapping frequency and change ratio between them. Term cluster mapping frequency meant the percent of gene products in the input gene list mapped to the specific term. Background mapping frequency presented the percent of gene products in the whole genome gene list mapped to the specific term. Detailed gene lists mapping to a specific term in the input gene list were also available for the user. Meanwhile, the table also offered the relationship and location network diagram, if the user clicked the corresponding button. As shown at the lower right in Figure 5, the hierarchy network diagram contained term nodes (node color meant different change ratio values, node color was close to red if high value and blue if low value). Overall, the platform implemented a GO annotation enrichment analysis by simple operations, detailed term mapping information, relationship, and location information to assist the user to further narrow the scope of target genes.

Discussion

The strengths of our online analysis platform compared with other tools are its easiness to operate for biomedical researchers without bioinformatics training and directly provides potential driver genes as research objects. In general, this platform was designed to visualize genomic alterations of a single gene among tumor samples in TCGA and apply novel mining algorithms to ranking genes to offer further study for cancer researchers. In addition, we expanded to utilize gene lists to map to GO terms to narrow down the scope of target genes. The GO term relationships and locations were visualized for the user to help generate hypotheses to explain tumor development.

Although only two cancer types were used in this study, it is easy to extend the platform to more cancer types. We plan to incorporate more genomic data types (e.g., miRNA/mRNA sequencing) and more detailed query conditions according to different clinical data (such as age, sex, and smoking status) to help the user explore the TCGA in a different way. Furthermore, we will add more driver gene mining algorithms and more adaptive GO Annotation analyses.

Conclusion

We developed a web-based genomic analysis platform as a bridge between big data and biomedical researchers. Researchers could conveniently use the platform to query gene-related genomic alteration information in a user-friendly way. Through embedded driver gene mining algorithms and GO annotation analysis, cancer researchers with limited bioinformatics resources can easily explore the TCGA data. This publicly accessible platform will help cancer researchers generate and test their hypotheses on TCGA data more easily, better facilitating translational research.

Acknowledgments

This research was financially supported by the National High-tech R&D Program of China (2012AA02A601) and National Natural Science Foundation of China (30900329).

References

[1] Gao J, Ciriello G, Sander C, Schultz N. Collection, integration and analysis of cancer genomic profiles: from data to insight. Current opinion in genetics & development. 2014;24:92-8.

[2] Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia MM, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerod A, Lee MTM, Shen CY, Tee BTK, Huimin BW, Broeks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, Van 'Tveer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis JS, Aparicio SAJR, Salomon AV, Borresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, Stratton MR, Osbreac. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486(7403):400-+.

[3] Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W, Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, Faham M, Seshagiri S. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010;466(7308):869-73.

[4] Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

[5] Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.

[6] Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43-9.

[7] Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507(7492):315-22.

[8] Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67-73.

[9] Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 2013;14(2):178-92.

[10] Gao Y, Chen X, Wang J, Shangguan S, Dai Y, Zhang T, Liu J. A novel approach for copy number variation analysis by combining multiplex PCR with matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Journal of biotechnology. 2013;166(1-2):6-11.

[11] Kuiper RP, Ligtenberg MJ, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. Current opinion in genetics & development. 2010;20(3):282-9.

[12] Trifonov V, Pasqualucci L, Dalla Favera R, Rabadan R. MutComFocal: an integrative approach to identifying recurrent and focal genomic alterations in tumor samples. BMC systems biology. 2013;7:25.

[13] Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics. 2004;20(18):3710-5.

[14] Spinelli L, Gambette P, Chapple CE, Robisson B, Baudot A, Garreta H, Tichit L, Guenoche A, Brun C. Clust&See: a Cytoscape plugin for the identification, visualization and manipulation of network clusters. Bio Systems. 2013;113(2):91-5.

Address for correspondence

Dr. Haomin Li,Zhouyiqing Building #510, Yuquan Campus, ZheJiang University, Hangzhou, Zhejiang 310027, Tel: 086+57187952383, Email: hmli@zju.edu.cn

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-663

Conceptual Knowledge Discovery in Databases for Drug Combinations Predictions in Malignant Melanoma

Kelly Regan[a], Satyajeet Raje[a], Cartik Saravanamuthu[a], Philip R.O. Payne[a]

[a] *Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA*

Abstract

*The worldwide incidence of melanoma is rising faster than any other cancer, and prognosis for patients with metastatic disease is poor. Current targeted therapies are limited in their durability and/or effect size in certain patient populations due to acquired mechanisms of resistance. Thus, the development of synergistic combinatorial treatment regimens holds great promise to improve patient outcomes. We have previously shown that a model for in-silico knowledge discovery, Translational Ontology-anchored Knowledge Discovery Engine (TOKEn), is able to generate valid relationships between bimolecular and clinical phenotypes. In this study, w*e have aggregated observational and canonical knowledge consisting of *melanoma-related biomolecular entities and targeted therapeutics in a computationally tractable model. We demonstrate here that the explicit linkage of therapeutic modalities with* biomolecular underpinnings of melanoma utilizing the TOKEn pipeline *yield a set of informed relationships that have the potential to generate combination therapy strategies.*

Keywords:

Malignant Melanoma; Knowledgebases; Combination Drug Therapy.

Introduction

Melanoma is the most deadly form of skin cancer, accounting for nearly 10,000 deaths in the United States in 2014. The incidence of melanoma is rising faster than any other cancer in the U.S., and there were over 76,000 new cases diagnosed in 2014 [1]. The death rate for melanoma patients in the U.S. has remained stagnant for the past 20 years, and less than 20% of patients have shown responses to traditional chemotherapeutic therapies [2, 3]. Cancer-driving *BRAF* mutations (V600E/K) are found in 40-60% of melanoma patient tumors, and BRAF-inhibitor agents, dabrafenib and vemurafenib, have extended median patient survival by 5-6 months [4]. Despite recent advances in targeted therapies, drug resistance remains a significant challenge for melanoma patients. Thus, further work to discover drugs that act synergistically with existing therapies and decrease drug resistance is desirable.

Traditional bench-based approaches for discovering synergistic drug combinations, including high-throughput drug screening, are costly and inefficient [5, 6]. It is estimated that an average of 1 billion dollars and 15-20 years is needed to bring a new drug from the bench to the bedside [7]. Further, 52% of drugs fail during development in phase 1 clinical trials, and only 25% of compounds that enter phase 2 proceed into full phase 3 clinical studies [8]. Biomedical informatics methods may offer more efficient and efficacious approaches for identifying synergistic drug combinations. Several computational approaches to optimize the drug discovery process have been proposed that involve modeling of structural, biochemical and biophysical properties [9].

In this study, we aim to computationally identify possible drug combinations to act synergistically with BRAF inhibitor therapies using a knowledge-anchored approach. The use of Conceptual Knowledge Discovery in Databases (CKDD) methods provides a potential means to accelerate hypothesis generation and recapitulation of known relationships between combinations of database entities. We have previously shown that a model for *in-silico* knowledge discovery, Translational Ontology-anchored Knowledge Discovery Engine (TOKEn), is able to generate valid relationships between bimolecular and clinical phenotypes in the context of large-scale, chronic lymphocytic leukemia datasets [10].

Knowledge discovery in databases represents a type of conceptual knowledge engineering method used to characterize relationships among distinct elements contained within a database [11]. Domain-specific knowledge collections, such as ontologies, are commonly used during knowledge discovery to augment meta-data contained in the targeted database schema. This overall approach is the basis for constructive induction, a type of knowledge discovery in databases (Figure 1). The constructive induction process generates conceptual knowledge constructs, otherwise referred to as induced "facts," that are defined by data elements and the semantic relationships that link them. Resulting conceptual knowledge constructs may be used to generate potential hypotheses about relationships between distinct data elements. Previous evaluation of the TOKEn method demonstrated its validity and "meaningfulness" according to domain experts [10]. Here we present the first application of TOKEn aimed at identifying drug combinations in malignant melanoma.

Methods

The TOKEn workflow has been previously described [10]. The overall workflow specifically applied in this study is shown in Figure 2. We obtained 42 FDA-approved and investigational "melanoma" drugs from DrugBank (version 4.1) [12]. DrugBank is a comprehensive database that includes chemical, pharmacological and pharmaceutical drug information as well as sequence, structure and pathway information regarding drug targets into more than 200 data fields per therapeutic agent. Relevant data fields pertaining to biomolecular foundations of drug action were selected, including "description," "mechanism of action," "pharmacodynamics" and "targets."

We developed an automated method to map selected DrugBank database fields containing free text to concepts within the Unified Medical Language System (UMLS), and selected those concepts belonging to the NCI, SNOMED-CT, MSH and GO ontologies due to their broad coverage,

including concepts related to drug features and actions. Similarly, a set of semantic types was heuristically defined to generate hypotheses targeted to drugs. Mapped entities were subsequently reviewed manually for accuracy and relevancy for mechanistic underpinnings of therapeutic agents. We obtained UMLS Metathesaurus associations from the previously curated set including parent, child and semantic relationships that were refined by subject matter experts to filter those relationships to be most meaningful for relating biomolecular and phenotypic concepts [10]. We determined that these heuristics generated in the original TOKEn study to be sufficiently generalizable for our purposes. We set search space optimization controls for the constructive induction method by calculating the shortest path depth-from-root of the ontology concepts selected and used them to annotate concepts as an indicator of concept granularity.

The UMLS MRHIER source file indexes all unique hierarchical paths determined by the source vocabulary as strings of distinct atoms from a particular concept to the UMLS root concept. Using this file, the minimum distance to the root was calculated for each UMLS concept corresponding to the source vocabularies. For each concept unique identifier (CUI), we set the 'minimum distance to the root' equal to the minimum number of elements in the corresponding path-to-root fields. The average depth of the ontology concepts that were mapped from the initial DrugBank data elements was found to be 4 "steps" from the UMLS root. We generated induced "facts" (Figure 1) using the graph-theoretic constructive induction algorithm previously described [10]. Traversal paths for drug combinations initiated at concepts associated with BRAF inhibitor drugs (vemurafenib, dabrafenib, PLX-4032) and terminated at those associated with the remaining 39 non-BRAF inhibitor drugs in the set. The algorithm avoids cycles by preventing the inclusion of duplicate concepts within a single traversal path. We constrained all concepts included in the induced "facts" to be at a depth equal to or greater than the minimum of the initial and terminal concepts. Pairs include direct relationships between drug-related concepts, while triples and quadruples include 1 and 2 intermediate concepts, respectively.

In order to prioritize drug relationships generated via the TOKEn method, we incorporated a novel ranking method according to their relatedness to melanoma pathogenesis. We obtained 663 "melanoma" concepts within the NCI Thesaurus (version 14.07d), and of those we used 221 concepts related to biomolecular properties of the disease. For example, we excluded tissue-level diagnoses or other high-level disease terms (uveal melanoma, acral melanoma, etc.). We further applied the TOKEn method to generate relationships that initiated with these NCI melanoma-associated concepts and terminated with concepts derived from the 42 DrugBank melanoma-associated therapies.

In order to rank BRAF inhibitor and non-BRAF inhibitor drug pairs generated via TOKEn, we calculated a Drug Combination Score (DCS) using the sum of two metrics for non-BRAF inhibitor drugs. Since BRAF inhibitors are expected to have the same degree of relatedness to melanoma pathogenesis, dabrafenib, vemurafenib and PLX-4032 were weighted equally. For each non-BRAF inhibitor drug, we summed: the Overlap score, the number of concept unique identifiers that mapped directly to drug concepts and those intermediate concepts derived from the induced facts between drugs that overlapped with the melanoma-associated set of CUIs; and the Distance score, the number of induced fact relationships between drug and melanoma concepts that were also weighted to the number of hops between initial and terminal concepts, where proximal relationships (i.e. fewer hops) were given a higher weight. Resulting Drug Combination Scores (DCS) were rank-ordered from highest to lowest for each non-BRAF inhibitor drug and final paths to concepts BRAF inhibitor. Log-2 transformed values for scores are reported.

Results

Concept identification and constructive induction

The 42 melanoma drugs indexed in DrugBank were mapped to UMLS concepts. Following manual review, a total of 495 drug-related UMLS concepts were identified for this study. The mean and median number of unique concepts per drug were 11.8 and 10.5, respectively. The BRAF inhibitor drugs (n=3) and non-BRAF inhibitor drugs (n=39) mapped to 23 and 300 unique concepts, respectively. The total numbers of induced facts in this study are listed in Table 1. For the total number of pairs, triples, quadruples and quintuplets, 28, 187, 6,469, and 196,284 concepts were anchored to a BRAF inhibitor. A total of 202,968 induced facts between BRAF inhibitor and non-BRAF inhibitor drugs were generated. In Table 2, examples of induced traversal paths between BRAF inhibitor drugs (vemurafenib, dabrafenib, PLX-4032) and non-BRAF inhibitor drugs are shown.

Drug combination scoring of induced facts

The FDA-approved anti-melanoma drug combination of trametinib and dabrafenib is evidenced here (Table 2) by the recognized relationship between the MAP2K1 and BRAF proteins in the MAPK signaling pathway and phosphorylation of its constituent proteins in melanoma tumors [13]. Furthermore, the combination regimen of trametinib and dabrafenib was recently approved by the FDA for use in melanoma patients with BRAF V600E or V600K mutations. Although PI-88 is an investigational drug not currently approved by the FDA, we show here evidence that it may support inhibition of tumor angiogenesis in combination with other BRAF inhibitors.

We implemented the DCS scoring metric to rank proposed BRAF inhibitor and non-BRAF inhibitor drug combination pairs predicted by the TOKEn algorithm (Table 3). The number of unique mapped concepts for the BRAF inhibitors dabrafenib, PLX-4032 and vemurafenib were 16, 3, and 14, respectively. Importantly, all three BRAF inhibitors shared the common set of concepts "BRAF gene," "Proto-Oncogene Proteins B-raf," and "Phosphotransferases" that were identified as initial concepts in all induced relationships that terminated with those associated with non-BRAF inhibitor drugs. Due to this congruency among concepts and the common therapeutic action of inhibiting BRAF protein activity, BRAF inhibitors were weighted equally in our scoring algorithm.

The Distance score component of the DCS was calculated over 69,856 induced facts between non-BRAF inhibitor and melanoma concepts. The values of non-zero log-2 transformed DCS ranged from 4.25 (AS1409) to 28.83 (AGRO100). In principle, the Overlap score emphasizes the direct relationships between drug and melanoma concepts (e.g. drug targets representing melanoma genes), with the tradeoff of possibly severely limiting potential drug-disease connections. Conversely, the Distance score emphasizes indirect relationships, or induced facts, between drug and melanoma concepts. Of note, the Overlap scores and Distance scores were significantly correlated among all non-BRAF inhibitor drugs (Pearson = 0.46, $p<0.0031$). Thus, these component metrics can be viewed as complementary when used to rank

drug combination predictions. Others have shown that distance-based metrics can be used to determine the similarity between drugs based on conceptual knowledge for drug repurposing applications [14-17]. Future work will assess the individual contributions of these metrics in validated drub combination relationships.

Discussion

In this study, we present the first application of the TOKEn method for *in-silico* knowledge discovery to database-derived melanoma drug-disease relationships. We developed a novel ranking metric for TOKEn-generated hypotheses for BRAF inhibitor (n=3) and non-BRAF inhibitor (n=39) drug combinations in melanoma. We found 202,968 unique relationships linking BRAF inhibitor drugs to 30 out of 39 non-BRAF inhibitor melanoma drugs that were indexed in DrugBank. AGRO100 was the highest ranked non-BRAF inhibitor drug for use in combination with BRAF inhibitors. AGRO100 is an experimental anticancer agent that acts as an aptamer and inhibits nucleolin, a protein that is uniquely expressed on the surface of tumors cells.Results from a phase I study of AGRO100 in patients with advanced cancers showed that half of the patients enrolled had stable disease with no toxic effects observed in any patient [18]. Further clinical development of AGRO100 resulted in the modified AS1411 agent, which has shown subsequent success in clinical applications [19, 20]. Interestingly, AS1411 was shown to reduce levels of a specific subset of miRNAs through its actions on nucleolin, including miR-21, miR-221, miR-222 and miR-103. These miRNAs are causally involved in breast cancer initiation, progression and drug resistance, and are also well known melanoma-associated miRNAs. Of note, all of the interferon-based therapies, with the exception of natural alpha interferon, were highly ranked according to their DCSs (rank 2-4).

Recently, combinations of BRAF inhibitors and immunotherapies have shown improved efficacy treating melanoma tumors and are being investigated in clinical trials [21]. Trametinib, a MEK inhibitor, is part of an FDA-approved combination with dabrafenib, and was ranked 7th according to DCS and 1st according to the overlap score alone. AZD-8330 is an experimental MEK inhibitor, and was ranked 5th according to DCS and 2nd according to the overlap score alone. This application of the TOKEn method has made improvements over the initial iteration by incorporating object-oriented programming and a system by which to prioritize induced facts.

Our study is currently limited by constraining the TOKEn method to canonical knowledge regarding melanoma concepts. The limited information contained within structured databases such as DrugBank may have prohibitively reduced our search space, and may have accounted for failing to recover relationships for 9 out of 39 non-BRAF inhibitor drugs. Furthermore, by limiting this study to known melanoma drugs indexed within DrugBank, we excluded possibilities to evaluate other existing drugs and small molecules that are not currently indicated or investigated in melanoma. Future work will include increasing our search space to other databases targeted to drug and melanoma information, as well as extracting data directly from the literature. We will also conduct future work to validate our findings and approach, including evaluating semantic similarity measures among mapped concepts, using graph-based network modeling methods and formal subject matter expert review.

Conclusion

We have demonstrated that a model for in-silico knowledge discovery, Translational Ontology-anchored Knowledge Discovery Engine (TOKEn), is able to generate valid relationships between drug and biomolecular phenotypes in the context of malignant melanoma.

Table 1– Summary of transitive paths generated at a search depth control of 5 and a distance from root of 4.

Number of concepts in induced "facts"	**2**	**3**	**4**	**5**
Number of unique relationships	5,940	103,540	4,789,356	100,289,621

Table 2– Examples of predicted "facts" connecting distinct drugs via triplets.

Relationship pattern	Conceptual knowledge constructs
Trametinib → Dabrafenib	MAP2K1 protein – [gene plays roles in biological process] – Serine/Threonine Phosphorylation – [process involves gene] – BRAF gene
PI-88 → Vemurafenib	FGF1 gene – [gene plays role in process] – Angiogenic process

Table 3– Top ten ranked non-BRAF inhibtor drugs hypothesized for use in combination with BRAF inhibitor drugs. DCS = Drug Combination Score. Reported DCS, Overlap and Distance scores are log-2 transformed values

Non-BRAF inhibitor drug	**DCS**	**Overlap score**	**Distance score**
AGRO100	*28.83*	*14.46*	*14.37*
Peginterferon-alfa-2a	*28.66*	*13.93*	*14.74*
Interferon_alfacon-1	*28.45*	*13.87*	*14.58*
Inteferon_Alfa-2b	*25.82*	*11.24*	*14.58*
AZD-8330	*25.77*	*14.68*	*11.10*
Trabectedin	*25.33*	*11.92*	*13.41*
Trametinib	*25.02*	*14.99*	*10.04*
ZEN-012	*23.09*	*11.40*	*11.70*
PI-88	*23.01*	*11.09*	*11.93*
ABT-510	*22.31*	*10.99*	*11.33*

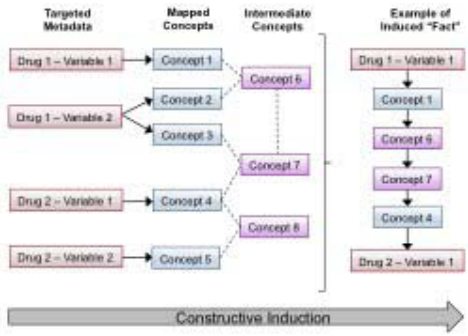


Figure 1 – Constructive induction of conceptual facts between distinct drugs. Mapping between database elements of targeted metadata to corresponding ontology concepts are utilized to induce "facts" among database elements, in this case, distinct drugs. Concepts 6 and 7 represent intermediate concepts not mapped to an original drug database element that define a higher-order transitive path that begins and terminates with drug database elements.

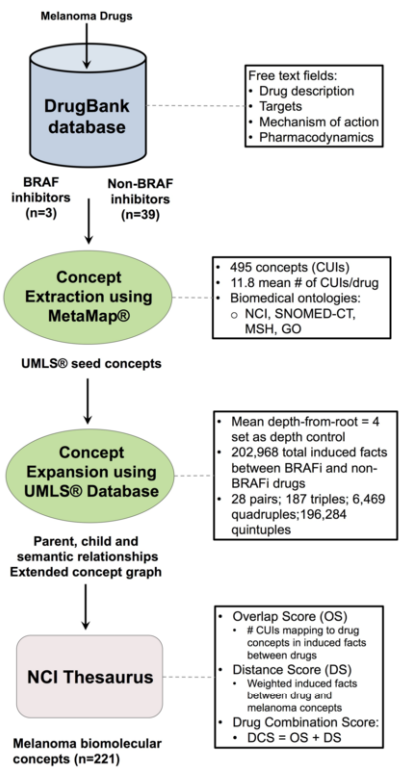


Figure 2 – Overview of TOKEn and DCS workflow.

References

[1] Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, and Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Bethesda, MD,* http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER website, April 2014.

[2] Chapman PB, Einhorn LH, Meyers ML, Saxman S, Destro AN, Panageas KS, Begg CB, Agarwala SS, Schuchter LM, Ernstoff MS, Houghton AN, and Kirkwood JM. Phase III multicenter randomized trial of the Dartmouth regimen versus dacarbazine in patients with metastatic melanoma. *J Clin Oncol* 1999: 17 (9): 2745-51.

[3] Kim KB, Sosman JA, Fruehauf JP, Linette GP, Markovic SN, McDermott DF, Weber JS, Nguyen H, Cheverton P, Chen D, Peterson AC, Carson WE 3rd, and O'Day SJ. BEAM: a randomized phase II study evaluating the activity of bevacizumab in combination with carboplatin plus paclitaxel in patients with previously untreated advanced melanoma. J Clin Oncol 2012: 30 (1): 34-41.

[4] Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, and McArthur GA. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011: 364 (26): 2507-16.

[5] Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, Price ER, Serbedzija G, Zimmermann GR, Foley MA, Stockwell BR, and Keith CT. Systematic discovery of multicomponent therapeutics. Proc Natl Acad Sci U S A 2003: 100 (13): 7977-82.

[6] Zhang L, Yan K, Zhang Y, Huang R, Bian J, Zheng C, Sun H, Chen Z, Sun N, An R, Min F, Zhao W, Zhuo Y, You J, Song Y, Yu Z, Liu Z, Yang K, Gao H, Dai H, Zhang X, Wang J, Fu C, Pei G, Liu J, Zhang S, Goodfellow M, Jiang Y, Kuai J, Zhou G, and Chen X. High-throughput synergy screening identifies microbial metabolites as combination agents for the treatment of fungal infections. Proc Natl Acad Sci U S A 2007: 104 (11): 4606-4611.

[7] Adams CP and Brantner VV. Estimating the cost of new drug development: is it really 802 million dollars? Health Aff (Millwood) 2006: 25 (2): 420-8.

[8] Bunnage ME. Getting pharmaceutical R&D back on target. Nat chem biol 2011: 7 (6): 335-9.

[9] Sahoo MB, Dinda SC, Ravi Kumar BVV, and Panda JR. Computational approaches for drug design and discovery process. Current Pharma Research 2012: 2 (3): 600-11.

[10] Payne PR, Borlawsky TB, Lele O, James S, and Greaves AW. The TOKEn project: knowledge synthesis for in silico science. J Am Med Inform Assoc 2011: 18 (1 Suppl): i125-31.

[11] Payne PR, Mendonca EA, Johnson SB, and Starren JB. Conceptual knowledge acquisition in biomedicine: A methodological review. J Biomed Inform 2007: 40 (5): 582-602.

[12] Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, and Wishart DS. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014: 42: D1091-7.

[13]Menzies AM and Long GV. Dabrafenib and trametinib, alone and in combination for BRAF-mutant metastatic melanoma. Clin Cancer Res 2014: 20 (8): 2035-43.

[14]Gramatica R, Di Matteo T, Giorgetti S, and Barbiani M,Graph theory enables drug repurposing – how a mathematical model can drive hidden mechanisms of action. PLoS One 2014: 9 (1): e84912.

[15]Andronis C, Sharma A, Virvilis V, Deftereos S, and Persidis A. Lifterature mining, ontologies and information visualization for drug repurposing. Brief Bioinform 2011: 12 (4): 357-368.

[16]Wang L, Liu H, Chute CG, and Zhu Q. Cancer based pharmacogenomics network supported with scientific evidences: from the view of drug repurposing. BioData Mining 2015: 8 (9). Doi: 10.1186/s13040-015-0042-8.

[17]Bisgin H, Liu Z, Kelly R, Fang H, Xu X, and Tong W. Investigating drug repositioning opportunities in FDA drug labels through topic modeling. BMC Bioinformatics 2012: 13 (Suppl 15): S6.

[18]Laber DA, Sharma VR, Bhupalam L, Taft B, Hendler FJ, and Barnhart KM. Update on the first phase I study of AGRO100 in advanced cancer. J Clin Oncol, 2005 ASCO Annual Meeting Proceedings 2005: 23 (16S): 3064.

Address for correspondence

Kelly.Regan@osumc.edu

[19] Rizzieri D, Stockert-Goldstein K, Wei A, Herzig RH, Erlandsson F, and Stuart RK. Long-term outcomes of responders in a randomized, controlled phase II trial of aptamer AS1411 in AML. *J Clin Oncol, 2010 ASCO Annual Meeting Abstracts* 2010: 28 (15S): 6557

[20]Pichiorri F, Palmieri D, De Luca L, Consiglio J, You J, Rocci A, Talabere T, Piovan C, Lagana A, Cascione L, Guan J, Gasparini P, Balatti V, Nuovo G, Coppola V, Hofmeister CC, Marcucci G, Byrd JC, Volinia S, Shapiro CL, Freitas MA, Croce CM. In vivo NCL targeting affects.

[21]Hu-Lieskovan S, Mok S, Moreno BH, Tsoi J, Robert L, Goedert L, Pinheiro EM, Koya RC, Graeber TG, Comin-Anduix B, and Ribas A. Improved antitumor activity of immunotherapy with BRAF and MEK inhibitors in BRAFV600E melanoma. Sci Transl Med 2015: 7 (279): 279ra41

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-668

Managing OMICS-Data: Considerations for the Design of a Clinical Research IT-Infrastructure

Nadine Umbach[a,b,c], Benjamin Löhnhardt[b,d], Ulrich Sax[a,b]

[a] *Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany*
[b] *GenoPerspektiv Consortium, Germany*
[c] *DZHK (German Centre for Cardiovascular Research), Partner Site Göttingen, Göttingen, Germany*
[d] *Operational Division of Information Technology, University Medical Center Göttingen, Göttingen, Germany*

Abstract

Biomarker-based translational research enables deep insight into cellular processes and human diseases. As a result, high-throughput technologies promulgating a fast and cost-effective generation of data are widely used to advance our understanding in the molecular background of individuals. However, the increasing volume and complexity of data increases the need for sustainable infrastructures and state-of-the-art tools allowing management, analysis, and integration of OMICS data. To address these challenges, we have performed site visits of core facilities with a focus on high-throughput technologies to explore their (IT) infrastructure, organizational aspects, and data management strategies. Different stakeholders were interviewed regarding requirements and needs for dealing with high-throughput data. We have identified four different fields of action: (1) the interface from biorepositories to service providers of high-throughput technologies, (2) aspects within services providers, (3) the interface from service providers to bioinformatical analysis, and (4) organizational and other aspects. For each field, recommendations and strategies were developed for implementation of a seamless pipeline from biorepositories to highly specialized high-throughput laboratories including the sustainable management and integration of OMICS data.

Keywords:

OMICS; clinical research infrastructure; metadata standards; data management.

Introduction

In personalized medicine, biomarker-based research is widely applied for translating biological knowledge into diagnostic, predictive, and therapeutic application (taking differences between individuals into account). Here, a variety of molecular high-throughput technologies are used to advance our understanding in cellular processes and human diseases. This includes analysis of the genome, transcriptome, epigenome, proteome, and metabolome. In colloquial language, these disciplines end with the suffix “omics”. Widely applied methods and technologies in the OMICS field are next-generation sequencing (NGS), microarrays, and liquid chromatography in combination with mass spectrometry. They all share a high degree in miniaturization, automatization, and parallelization.

Usually, high-throughput platforms are operated in centralized core facilities (CF) of scientific research institutions or health centers and have a focus on specialized disciplines (e.g., genomics or proteomics). Beside the efficient and effective provision of methodological skills, the operation of highly specialized equipment, and the bioinformatic data analysis; CF offer comprehensive consulting in project design, further development and maintenance of applications, and tutorials for the transfer of expert knowledge. In addition to CF, high-throughput platforms and profound knowledge can also be found in biomedical or natural sciences departments. In both cases, services are available to research groups affiliated to the institution and external cooperation partners, in compliance with dedicated concepts, rules, and conditions. Over the last years, technological improvements, significant cost reductions, and the faster analysis of biomolecules were the main impulses driving the field from research to clinical application. Here, the main focus is on the analysis of disease-related genes generally known as panels.

However, the growing popularity of OMICS also leads to big challenges [1-3]. The increasing volume of huge and highly complex data sets pose greater requirements on data management (including integration, analysis, archiving, and provision of data), server sizing, and computing power. The traceability and reproducibility of data also require comprehensive and standardized annotation of processes, equipment, and tools. Also, given the specific requirements on clinical data in terms of quality, priority, and validity, separate OMICS infrastructures are required for clinical context. In order to have a sustainable and long-lasting infrastructure, general concepts for data management and infrastructure organization are required. Here, we present strategies and recommendations for the design of a generic clinical research infrastructure for OMICS data; and an overview about tools for their management and analysis to accelerate translational research.

Materials and Methods

To address the current situation and needs of suppliers and users in terms of infrastructure organization and data management, and to develop strategies and recommendations for improved dealing with OMICS data; the following four steps were performed:

Design of the questionnaire

A structured interview guide was designed with the intention to address one issue per section and to analyze the current situation, challenges, and possible solutions. Identified topics were: (1) general questions concerning the CF and the local environment, (2) infrastructure organization, (3) offered services including their general regulations and conditions, (4) available equipment, (5) data management including concepts for data annotation, integration, archiving, and provision, and (6) bioinformatical issues. The interview guide was used as an

orientation guide for the interviews with experts in the field and for site visits of CF of biomedical and natural sciences departments with strong expertise in OMICS.

Identification of users and conduction of qualitative interviews

Four target groups were identified: (1) heads and managers of OMICS core facilities, (2) researchers and clinicians from the areas of human genetics, oncology / haematology, pathology, and pharmacology (users of OMICS technologies and data), (3) IT officers and computer scientists responsible for data management and IT infrastructure organization, and (4) bioinformaticians responsible for data analysis. In total, 15 interview partners were identified and contacted via E-Mail inviting them to participate in an interview. All interviews were conducted by the same two scientists to ensure standardized procedures. Interviews were recorded using a recording device and then transcribed. The transcription was restricted to the paraphrasing of statements made by the experts.

Identification of core facilities and conduction of site visits

In order to explore the processes, equipment with OMICS platforms and hardware, and organizational framework; site visits of CF were performed. Selection was based on two criteria: (1) focus on genomics, transcriptomics, proteomics, or metabolomics, and (2) affiliated with a scientific research institution or health center. Moreover, the selected facilities should represent a broad cross section of size, services, and equipment. In total, seven facilities were selected:

- the Microarray and Deep-Sequencing CF in Göttingen (Germany),
- the CF for Medical Biometry and Statistical Bioinformatics in Göttingen (Germany),
- the Institute for Clinical Molecular Biology in Kiel (Germany),
- the Functional Genomic Center in Zurich (Switzerland),
- the Interdisciplinary Center for Clinical Research in Leipzig (Germany),
- the Sequencing CF at the Max-Planck-Institute for Molecular Genetics in Berlin (Germany),
- the Institute of Experimental Genetics at the Helmholtz Zentrum in Munich (Germany),
- and the German Cancer Research Center and the University Hospital in Heidelberg (Germany).

Analysis of the results and conclusions drawn

Processes and results derived from the site visits were aggregated to a model infrastructure using Unified Modeling Language (UML) and Business Process Model and Notation (BPMN). Afterwards, requirements, needs, and challenges identified in expert's interviews were used to develop strategies and recommendations for improved management of OMICS data and infrastructure organization.

Results

As a result of our interviews, we find strong evidence, that due to today's high quality demands and the high degree of specialization we will see a consequential centralization of most biomarker service units towards high-throughput service providers. Considering the short model cycle and the high running expenses of sequencers only facilities with a high load factor and high quality interfaces from biobanks to the bioinformatic analysis groups will be sustainable.

Regarding the infrastructure, not only the measuring devices, but IT infrastructures are also cost drivers. A facility working to capacity will produce a similar amount of data as a Picture Archiving and Communication System (PACS) or a Pathology Information System in a hospital.

The service units mostly generate turnover from daily orders. As they are mostly only equipped with short-term storage, storing for longer time causes problems.

Therefore, the interfaces from the mostly vendor-specific file format from the measuring devices, the quality assurance in the primary analyses to the further steps outside the facility have to be taken into account.

Most relevant to our survey seems to be the pipeline from biorepositories and service providers of OMICS technologies as the data producing entities to the (secondary) bioinformatic analysis and data integration (Fig. 1). As the prognostic factor of OMICS data is very limited without further annotation, for example the phenotype data, structured anamnesis, and further clinical data have to be integrated prior to the overall analyses.

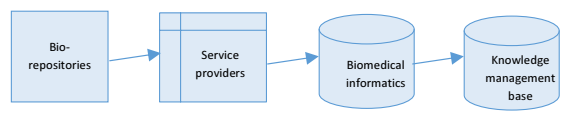


Figure 1 - Simplified pipeline from biorepositories to service providers of OMICS technologies (material logistics), transport of the resulting data to biomedical informatics units for integration of the corresponding data and further analyses. Furthermore, the results should contribute to a knowledge management base.

On the application level, we can distinguish four groups of software families along the high-throughput analysis pipeline.

While software for biorepositories has matured in recent years, the situation further down the pipeline seems to be much more complex. Within the service providers of OMICS technologies we find commercial software solutions with the corresponding analysis equipment [4]. Quite often this leads to vendor-specific file formats (e.g. bcl files for Illumina equipment) instead of standard formats like FASTQ [5, 6]. Unfortunately, there are still few standardized solutions for a workflow-supporting data management within most facilities from biorepositories to service providers of OMICS technologies and further down the pipeline to bioinformatic units. In consequence, this leads to individual solutions for the crucial transfer of analysis and result data from the sequencing facility to biostatisticians and bioinformaticians (iii).

For the integration of different data types (for example OMICS and clinical data) including the scripts for extracting, transforming, and loading the data (ETL), solutions are available [7] and discussed, evaluated, and further developed at many sites. As this is to be a larger scale problem, and initiatives like FAIRDOM[1] and research data alliance[2] are currently working on solutions. As a result, the openBIS [8] software seems promising to solve some workflow related problems.

In the last group we find some software packages that support the interface from data management to knowledge extraction

1 http://www.fair-dom.org
2 https://europe.rd-alliance.org/

and knowledge management (eTriks[3], bioxm[4] [9, 10], geneXplain[5]).

Discussion

In this paper, we identified four different fields of action: (1) the interface from biorepositories to providers of OMICS technologies, (2) aspects within providers of OMICS technologies, (3) the interface from providers of OMICS technologies to (bioinformatical) analytics, and (4) organizational and other aspects. The following guidance details each field of action:

Interface from biorepositories to providers of OMICS technologies

- While professional solutions documenting lab workflows, projects as well as management of biospecimens and their corresponding data are already widely spread and established in the domain of biobanking, there is a huge demand for the development and implementation of adequate IT solutions for managing OMICS labs.

Aspects within providers of OMICS technologies

- Adequate annotation schemes for the standardized and harmonized data acquisition from genomics, epigenomics, transcriptomics, proteomics, and metabolomics have to be developed and established.
- Already existing schemes for data annotation like the Minimum Information about a (Meta)Genome Sequence (MIGS) [11], the Sequence Read Archive (SRA) scheme of the European Nucleotide Archive, the Minimum Information About a Microarray Experiment (MIAME) [12], the Minimum Information About a Proteomics Experiment (MIAPE) [13, 14], and the Metabolomics Standards Initiative (MSI) have to be applied in practice, evaluated, and if needed, adopted.
- General overviews about existing laboratory equipment and a transparent (central-driven) procedure for the acquisition of new devices on a site are essential to achieve an efficient workload and to avoid unnecessary duplicate equipment acquisitions.
- For cost and quality reasons, it is to be expected over the medium term, that service providers in the OMICS field are becoming increasingly professionalized and centralized.
- Implementation and preservation of sustainable infrastructures cannot be done by project-related resources, but rather require the willingness of research institutions, institutions within the health system, and other national or international funding initiatives at the state level to make major investments and provide funds for operations.
- It has to be considered, that with falling prices for molecular high-throughput analysis, massive problems arise concerning long-term storage of data. To avoid the duplication of data at both providers' and customers' sites, processes and regulations are needed to clarify who has to preserve the data with respect to Good Scientific Practice (GSP) and Good Clinical Practice (GCP).
- Besides the raw data, long-term preservation has to contain quality scores, workload data from the labs and information about project design (including list of specimens and billing information). In the case of NGS analyses, raw data should be preserved as FASTQ, BAM or in vendor-specific formats from the laboratory devices (like bcl data format, which is generated by sequencers from Illumina). These vendor-specific formats allow for the deriving of all subsequent analysis data again.
- Many infrastructures for high-throughput data lack a standardized mechanism for data exchange between researchers, clinics, service providers, biomedical informaticians and biostatisticians. The definition and implementation of such interfaces are of major relevance. There is a need for action to achieve reproducibility and sustainable availability of data regarding GSP and GCP.

Interface from providers of OMICS technologies to analytics

- Integration, analysis, and interpretation of data require a profound knowledge about bioinformatic and biostatistic analysis as well as an understanding of the biological system context. However, the knowledge of researchers and physicians is oftentimes insufficient. To eliminate and avoid misunderstandings, the management, integration, and analysis of molecular high-throughput data have to be added to researchers' and physicians' curricula.
- A close collaboration between the OMICS fields and bioinformatics is necessary in the course of standardization, management, integration, and transfer of data and data models in systems biology.
- The lack of sustainable, site-independent solutions seems to be a widespread problem. International efforts are required in order to design such an infrastructure. Existing products like openBIS emerged from the Fairdom [8] project.

Organizational and other aspects

- For the sustainable operation of central service facilities, preservation and development of know-how as well as hands-on expertise are essential. Problems due to temporary mid-level academic positions have to be covered. To establish satisfactory expertise, these service facilities should also have a close connection to other related institutions.
- Existing infrastructure is predominantly research-oriented. Because of high requirements for patient care regarding quality, availability, and validity of high-throughput molecular data, separate structures are needed. Concrete and preferably generic concepts for the integration of OMICS analysis in clinical routine should be developed and implemented.
- Ethical and legal questions - especially for managing incidental findings and findings, which affect direct family members or derive from new scientific knowledge - have to be considered.

3 http://www.etriks.org/

4 http://www.biomax.com/products/bioxm-knowledge-management-environment/

5 http://genexplain.com/genexplain-platform-1

Conclusions

From literature, interviews, site visits, we can infer that there are high-quality solutions established within the professional high-throughput labs. Mostly, the vendor-software corresponding to the lab equipment is embedded in individual solutions for accepting and integrating data from the clients, individual interfaces for transferring the data to further analysis or back to the client.

This severely hampers efficient workflows. Urgent action is needed for the design of standardized interfaces between service providers, biostatisticians and users in order to adequately integrate data and reach the goal of reproducible, GSP and GCP-conforming data management in the OMICS arena.

The increased need in personalized medicine for management and exploration of OMICS and clinical data poses a big challenge for data management. Available solutions [7] are discussed, prototypically implemented, and evaluated at many sites. The further development and dissemination of such tools in the OMICS community will be crucial for translational research.

As the conversion of the hitherto scattered measuring equipment to sustainable, highly specialized high-throughput labs will continue, the seamless pipeline from biomaterial to integration, analysis, and knowledge management will be a key factor for success.

Although site visits and interviews in this report covers only German and Swiss institutions, the findings and recommendations are consistent with OMICS facilities in the UK, Paris (France), Pavia (Italy), and in Boston (US) reported by our colleagues within the US and European i2b2 academic user group.

In addition to the rather technical findings, quite some ethical and legal challenges are discussed nationally and internationally.

References

[1] Biesecker LG. The new world of clinical genomics. The Journal of clinical endocrinology and metabolism. 2012; 97(11):3912-4.

[2] Biesecker LG. Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. Genetics in medicine : official journal of the American College of Medical Genetics. 2012; 14(4):393-8.

[3] Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R. Next-generation sequencing in the clinic: are we ready? Nature reviews Genetics. 2012; 13(11):818-24.

[4] Berger B, Peng J, Singh M. Computational solutions for omics data. Nature reviews Genetics. 2013; 14(5):333-46.

[5] Bonfield JK, Mahoney MV. Compression of FASTQ and SAM format sequencing data. Plos One. 2013; 8(3):e59190.

[6] Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic acids research. 2010; 38(6):1767-71.

[7] Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. Briefings in bioinformatics. 2014; 16(2):280-90.

[8] Bauch A, Adamczyk I, Buczek P, Elmer FJ, Enimanev K, Glyzewski P, et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. BMC bioinformatics. 2011; 12:468.

[9] Cano I, Tenyi A, Schueller C, Wolff M, Huertas Miguelanez MM, Gomez-Cabrero D, et al. The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research. Journal of translational medicine. 2014; 12 Suppl 2:S6.

[10] Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, et al. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC systems biology. 2011; 5:38.

[11] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008; 26(5):541-7.

[12] Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001; 29(4):365-71.

[13] Taylor CF. Minimum reporting requirements for proteomics: a MIAPE primer. Proteomics. 2006; 6 Suppl 2:39–44.

[14] Taylor CF, Paton NW, Lilley KS, et al., The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol. 2007; 25(8):887–893.

Address for correspondence

nadine.umbach@med.uni-goettingen.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-672

Virtual Microscopy Large Slide Automated Acquisition: Error Analysis and Validation

Axel V. Mancino[a], Alyson Roger[a], Domingo L. Muscolo[b], German L. Farfalli[b], Pablo D. Roitman[c], Luis A. Aponte-Tinao[b], Lucas E. Ritacco[a], Federico E. Milano[ad]

[a] *Virtual Planning and Navigation Unit, Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina.*
[b] *Department of Orthopedic Surgery – Orthopedic Oncology, Hospital Italiano de Buenos Aires, Argentina.*
[c] *Department of Pathology, Hospital Italiano de Buenos Aires, Argentina.*
[d] *National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina.*

Abstract

The aim of this work is to assess and analyze the discrepancies introduced in the reconstruction of an entire tumoral bone slice from multiple field acquisitions of a large microscopy slide. The reconstruction tends to preserve the original structural information and its error is estimated by comparing the reconstructed images of eight samples against single pictures of these samples. This comparison is held using the Structural Similarity index. The measurements show that smaller samples yield better results. The detected errors are introduced by the insufficiently corrected optical distortion caused by the camera lens, which tends to accumulate along the sample. Nevertheless, the maximum error encountered does not exceed 0.39 mm, which is smaller than the maximum tolerable error for the intended application, stated in 1mm.

Keywords:
Microscopy; Confocal; XY-table; Stitching; Bone tumor; Histological Techniques; Error evaluation.

Introduction

In a primary bone tumor resection surgery, a surgical specimen is obtained following the guidance of a preoperative plan based on Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) studies [1,2,3]. In certain areas of the MRI images, however, the tumoral region limits cannot be well delineated by simple visual inspection.

Optical bright-field microscopy is a technique used in pathology to view samples and assess the classification of bone tumors [4,5]. The conventional practice in histopathology is to evaluate certain critical portions of the surgical specimen sample. Nevertheless, this practice does not enable establishing a direct correspondence between the microscopy images and preoperative MRI images, impeding the accurate classification of doubtful regions.

The main aim of this work is to enable the reconstruction of an entire bone slice sample from a surgical specimen, stitching together congruent microscopy pictures. Structural conservation is a major concern, due to the fact that the reconstructed image will be superimposed to its corresponding slice image in a volumetric magnetic resonance of the surgical specimen. Therefore an evaluation of the introduced error is addressed as well. The direct comparison between MRI image and microscopy images might lead to a new way of interpreting MRI images, and hence determine whether there exist tumoral tissues in the region under study.

Methods

Optical acquisition

In order to acquire the images, a Carl Zeiss microscope, Primo Star model (iLED Halogen/LED) was used, supplied with a Canon Powershot A640 camera, mounted through an array of coupling lenses. The total magnification rendered on the pictures corresponds to the objective magnification (40x) added to the camera analog zoom. This gives a total measured resolution of 1270 pixels per millimeter. The sample scanning is enabled by a XY-table designed with two step motors, providing a resolution of 200 steps per millimeter of movement (6.3 pixels per step). An 8-bits word UART serial communication between the controlling system, implemented in a PC running MATLAB® 2012b (The MathWorks, Inc., Natick, Massachusetts, United States), and the XY-table controller, implemented using a Texas Instruments® MSP430G2553 microcontroller completes the closed loop system.

Figure 1– Camera coupled to microscope, focusing a bone tissue sample disposed on the XY-table.

Sample scanning process

The scanning process is divided into two main steps:

- The control system, which takes consecutive pictures in the X axis direction of movement, generates a row

of the final image, and moves in the Y axis direction, leading to a new capture of pictures.

- The process of rows stitching.

Control system

Two consecutive pictures should have a transition region overlapping in a range from 40% to 60% of the capture area. Matching points between pictures are detected by searching common features with the Speeded-Up Robust Feature (SURF) algorithm [6]. Even though the movement is supposed to occur only in the X direction, there are displacements in Y due to the misalignment of the sample in the XY-table with the microscope-camera array. These displacements are compensated with consistent movements in the Y direction. During the search of features and execution of XY-table movements, a low-resolution video mode of the camera is used. Several high-resolution pictures are afterwards taken and their level of focus is measured using the Brenner gradient [7]. Once the best focused image is chosen, the process of stitching between consecutive images takes place. Afterwards, an optional blending step may be applied to match small differences in structure and luminescence in the transition zone between pictures.

Rows stitching

The rows previously obtained are then aligned one after the other, also using common features extracted with the SURF algorithm.

Optical compensations

Before starting the sampling process, luminescence compensation is performed: a sample-free RGB picture is taken (P), its mean is measured (μ), and the compensation RGB matrix (M) is calculated as follows:

$$Mij = \mu / Pij \quad (1)$$

Geometric compensation is also carried out, which attacks the fish-eye problem caused by the microscope-camera array lenses. This deformation consists on the stretching of the image increasing with the distance to its optical axes, and may cause image discontinuities after several pictures are attached together. In order to address this issue, an inverse deformation dependent on the square of the distance was applied [8]. This inverse deformation was estimated using pictures taken to a 0.1mm depth hemocytometer.

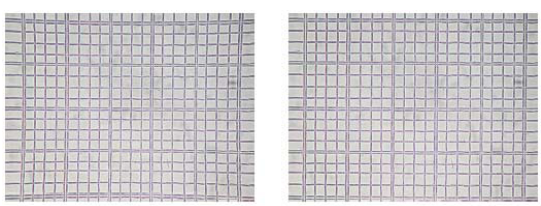


Figure 2– Fish-eye lens distortion geometrical correction applied to a 0.1mm depth hemocytometer picture.

Visualization

The software does not generate one final slide capture, but instead hundreds of small images, and a matrix that describes how to stitch them together. This design eases future work with higher magnification, since all that changes is the amount of pictures handled by the matrix.

Stitching error measurement protocol

Eight different samples were acquired, with sizes varying from 440 mm^2 to 1280 mm^2. In order to evaluate and determine the possible misalignments generated by the stitching process, a comparison between the final stitched image and a standard image must be performed. Thus, a not-amplified-by-the-microscope picture of the sample, taken using the same camera as in the stitching and corrected with the same geometrical factors, is used as a standard for comparison. Although it lacks the grade of definition of the stitched image, it serves as a pseudo ground truth to evaluate the level of preservation of image structure.

The method used for the quantification of this error uses the Structural Similarity (SSIM) index, combined with phase shifts between the images. The SSIM index ranks locally the similarity between portions of two images. The SSIM continuous range interval is -1 for worst case to 1 for perfect match. This index can be displayed as a gray level pixel map, giving a general idea of the amount of parity between images. The index per pixel is averaged to give an overall level of similarity, whose range limits are coincident with the previously mentioned.

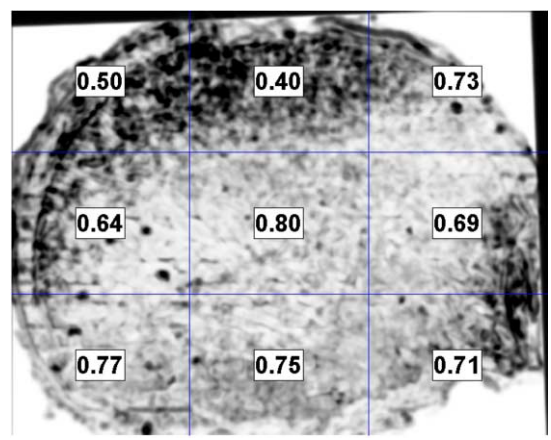


Figure 3– SSIM map displayed as (1+SSIM)/2 between the stitched image and the standard image, for Sample A. The map is divided in nine portions, where the corresponding index of each portion is labeled. Dark pixels mean bad correlation between images, while bright pixels correspond to good structural similarity.

The approach taken in this work to estimate the stitching error consists on registering the images, using rigid registration algorithms. Misalignment locations depend on registration process. If the correct registration and similarity congruency of a portion of the image leads to misalignments in other portions of the image, then a stitching error is identified. The stitching error in a part of an image may be compensated by adding translation offsets to the original registration. When moving one of the pictures a certain offset to the right, left, upwards, and downwards, SSIM map zones which originally appeared dark turn to white, meaning a good alignment in that part of the image, and vice versa.

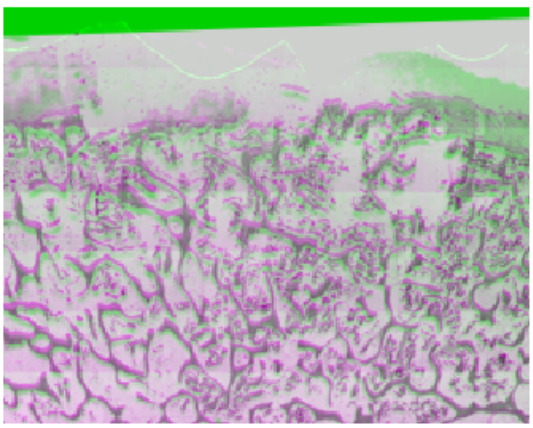


Figure 4– Portion of the image from Figure 3 (Sample A) with the lowest SSIM index (0.40), shown as the stitched image (magenta) superimposed in the standard image (green). As can be seen, a better registration can be achieved by moving the standard image upwards.

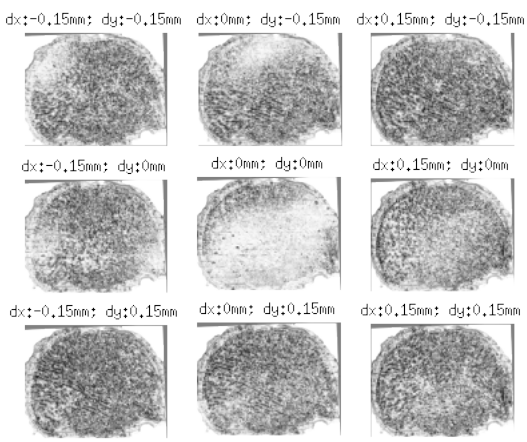
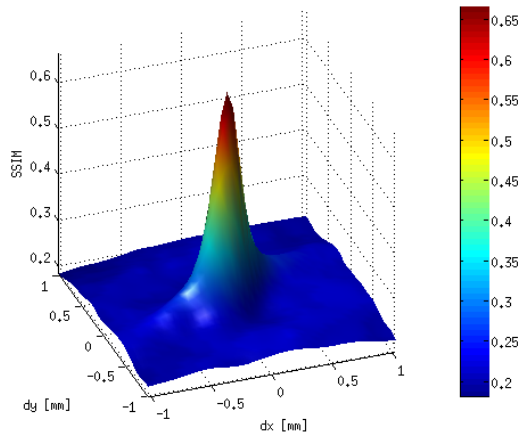


Figure 5– Gray level changes in the SSIM map representation of Sample A, as a function of offset. The center image represents the better registration found between images. Right (positive) and left (negative) shifts are labeled dx, while up (negative) and down (positive) are labeled dy.

Calculating the SSIM index after every shift leads to the construction of surfaces like the one shown in Figure 6. In order to determine the error introduced by the stitching process, it is necessary to evaluate the final image misalignments produced all along the sample, compared to the standard image. This error will be quantified as the minimum radius of displacement required, in millimeters, to reach the denominated noise floor of the SSIM index in the sample. Knowing the bell-like shape of the SSIM index evolution as a function of space and being proven that it is not affected by local maximums phenomena, the SSIM map is evaluated in the initially dark areas for different shifts in X and Y directions. Strictly, the noise floor is considered reached after subsequent movements in one direction do not improve the SSIM index, but worsen it. It should be highlighted that due to its accumulative behavior, the stitching error tends to grow with the distance. Though, it is expected that the major discrepancies will appear around the edges of the image (see Figure 3 and Figure 5), assuming a rigid registration towards the center of both images.



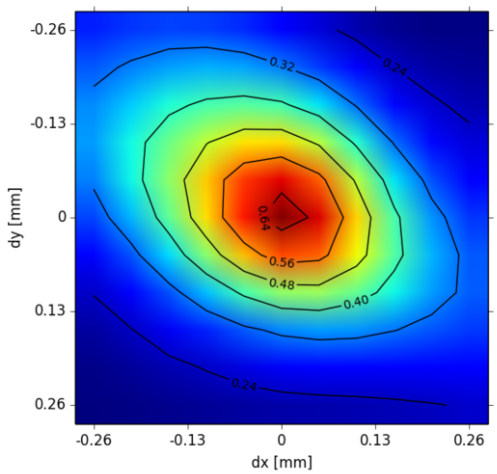


Figure 6– SSIM index as a function of image displacements (in X and Y) with respect to the better fixing registration employed in Sample A.

To evaluate an estimation of the general structural deviation of the stitched image proportions compared to the standard image ones, a Structural Deviation (SD) index is proposed for both X and Y maximun extensions (ΔX and ΔY correspondly):

$$SD_X = 100\ (\Delta Xstd - \Delta Xstitch) / \Delta Xstd \quad (2)$$

$$SD_Y = 100\ (\Delta Ystd - \Delta Ystitch) / \Delta Ystd \quad (3)$$

Results

The eight reconstructed samples (A to H) are displayed in Figure 7, and their error measurements are shown in Table 1.

Table 1– Error measurement of the samples stitched

S	Size [mm]	Total images	SD_X [%]	SD_Y [%]	SSIM index	Error [mm]
A	40 x 32	819	0.82	2.17	0.67	0.33
B	45 x 27	876	-3.27	-2.15	0.53	0.39
C	28 x 21	363	0.25	-2.08	0.64	0.28
D	20 x 22	266	-1.10	-1.44	0.73	0.10
E	26 x 22	364	-1.57	-0.13	0.67	0.26
F	29 x 24	392	-0.75	-2.77	0.65	0.22
G	23 x 23	326	-0.42	-1.45	0.75	0.16
H	32 x 22	378	-1.60	-0.35	0.63	0.20

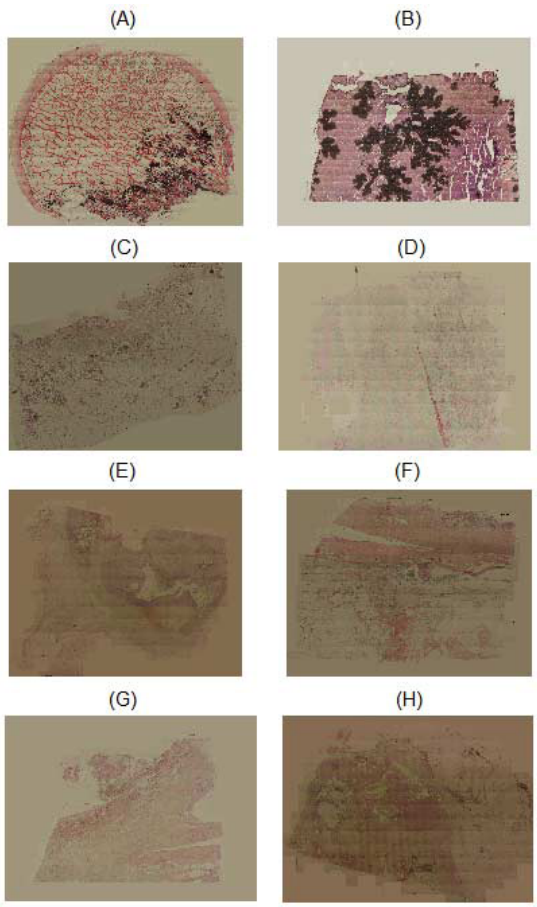


Figure 7– Images A to H showing the eight samples reconstructed using the stitching process.

Discussion

As it was shown in the results section, non-standard size (3 in x 1 in) samples could be reconstructed, with a 1270 pixels per millimeter resolution.

The results in Table 1 show an absolute maximum Structural Deviation of 3.27% in X and 2.77% in Y. In terms of stitching error, it can be seen that, in general, the higher the amount of pictures taken, the higher the error measured. Samples A and B appeared to be more deformed than the rest (0.33mm and 0.39mm), being both composed with more than three times as many images as Sample D (0.1 mm, minimum error). This deformation was caused by the accumulated stitching error along the entire sample, mostly due to the optical distortion produced by the camera lens used for acquisition, which is not perfectly corrected. In addition, this distortion was measured higher for thicker microscope slides (more than 4.5 mm thick), which were the ones used for Samples A and B. In order to lower the error caused by this effect, it is recommended to try to use thinner slides (1 mm thick) in all cases.

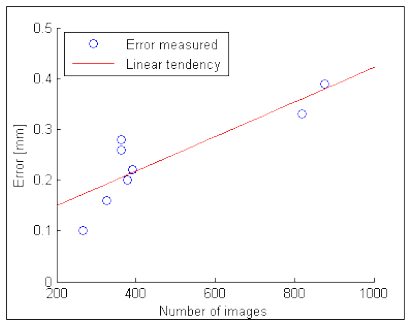


Figure 8– Error variation measured in the samples, as a function of the total number of images stitched.

Although the SSIM index is a relatively good index for assessing structural differences between images, in some cases it might perform poorly for the application presented in this work. Consider for example Figure 9. In spite of the fact that both images were taken with the same camera, the stitched image (B) presents much darker regions than the standard image (A), for the same tissue. This marked difference in the contrast of the images is interpreted by the SSIM index as a major structural difference, even though it is not.

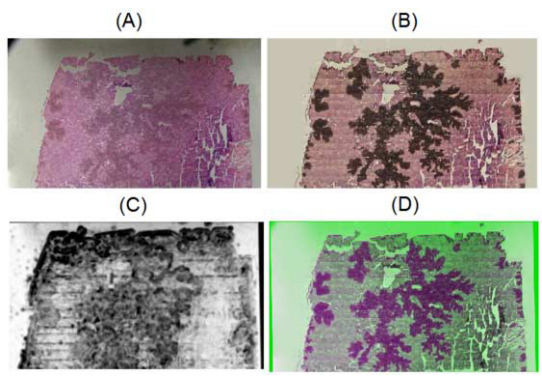


Figure 9– Although not a bone sample, this liver tissue sample clearly shows low rated areas due to contrast differences between images. (A) Standard image from sample B. (B) Stitched image from sample B. (C) SSIM map between standard image and stitched image from sample B. (D) Superposition of standard (green) and stitched (magenta) images from sample B. No significant structural differences can be seen.

Since the MRI resolution of the bone prior to becoming a surgical specimen is 1 mm, the same value will be taken as the maximum tolerable error in the complete stitched image.

It is considered essential the conservation of the original appearance of images. Although the blending between images

might lead to friendlier representations for the viewer, it could alter the specialist perception of the degree of cellular disorder appreciated, as shown in Figure 10. Hence, it is believed best to simply stitch the images one after the other, obviating any kind of image processing between them, both in the rows confection as in their alignments.

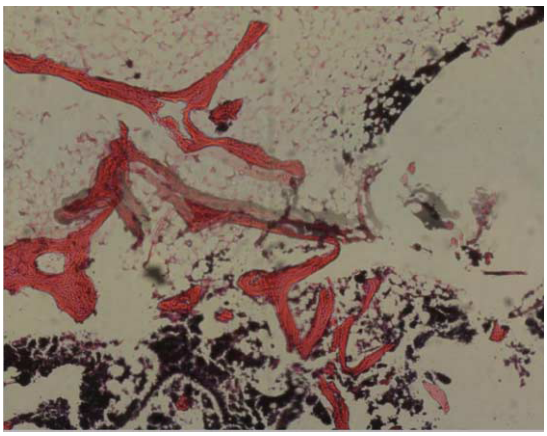


Figure 10– Stitching between two wrongly matched rows. The blending applied generates an unwanted ghost effect in the final image.

Conclusion

This work presents a new way to acquire histological images using a XY-table and a digital camera coupled to a microscope. Although being a relatively high time-demanding procedure (~30 seconds per image acquisition in average), it strongly lowers a Whole Slide Imaging Scanner budget. Measurement results presented show that the error committed is small enough for the suggested application, which aims, in the future, to help in the construction of a gold standard bone tumor samples repository.

Acknowledgments

We would like to acknowledge Luis H. Palacios for his collaboration in sample preparation and Horacio O. Milano for designing and building the XY-table hardware.

References

[1] Ritacco LE, Milano FE, Farfalli GL, Ayerza MA, Muscolo DL, Aponte-Tinao LA. Accuracy of 3-D planning and navigation in bone tumor resection. Orthopedics. 2013 Jul; 36(7): e942–50.

[2] Wooton-Gorges SL. MR Imaging of Primary Bone Tumors and Tumor- like Conditions in Children. Magn Reson Imaging Clin N Am. 2009 Aug; 17(3):469-87.

[3] Meyer JS, Nadel HR, Marina N, et al. Imaging guidelines for children with Ewing sarcoma and osteosarcoma: a report from the Children's Oncology Group Bone Tumor Committee. Pediatr Blood Cancer. 2008 Aug; 51(2):163–70.

[4] Nasir-Ud-Din, Ahmed A, Pervez S, Ahmed R, Kayani N. Chondroblastoma: A Clinico-pathological Analysis. J Coll Physicians Surg Pak. 2014 Dec; 24(12): 898-901.

[5] Hansford BG, Pytel P, Moore DD, Stacy GS. Osteoma of long bone: an expanding spectrum of imaging findings. Skeletal Radiol. 2014 Nov.

[6] Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 2008 Jun; 110(3): 346-359.

[7] Brenner JF, Dew BS, Horton JB, King T, Neurath PW, Selles WD. Automated microscope for cytologic research - preliminary evaluation. J. Histochem. Cytochem. 1976; 24:100–111.

[8] Ma L, Chen Y, Moore, KL. Flexible camera calibration using a new analytical radial undistortion formula with application to mobile robot localization. In IEEE International Symposium on Intelligent Control. 2003 Oct: 799 - 804

Address for correspondence

Department of Health Informatics,

Hospital Italiano de Buenos Aires.

Juan D. Peron 4190, C1181ACH Capital Federal, Buenos Aires

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-677

Do interoperable national information systems enhance availability of data to assess the effect of scale-up of HIV services on health workforce deployment in resource-limited countries?

Tom Oluoch[1], David Muturi[2], Rose Kiriinya[3], Anthony Waruru[1], Kevin Lanyo[3], Robert Nguni[4], James Ojwang[1], Keith P. Waters[5], Janise Richards[5]

[1] US Centers for Disease Control and Prevention/DGHA, Nairobi, Kenya
[2]USAID-Afya Info Project, Nairobi, Kenya
[3]CDC-Emory University Kenya Health Workforce Project, Nairobi, Kenya
[4]USAID-IntraHealth/Capacity Project, Nairobi, Kenya
[5]US Centers for Disease Control and Prevention/DGHA, Atlanta, GA, USA

Abstract

Sub-Saharan Africa (SSA) bears the heaviest burden of the HIV epidemic. Health workers play a critical role in the scale-up of HIV programs. SSA also has the weakest information and communication technology (ICT) infrastructure globally. Implementing interoperable national health information systems (HIS) is a challenge, even in developed countries. Countries in resource-limited settings have yet to demonstrate that interoperable systems can be achieved, and can improve quality of healthcare through enhanced data availability and use in the deployment of the health workforce. We established interoperable HIS integrating a Master Facility List (MFL), District Health Information Software (DHIS2), and Human Resources Information Systems (HRIS) through application programmers interfaces (API). We abstracted data on HIV care, health workers deployment, and health facilities geo-coordinates. Over 95% of data elements were exchanged between the MFL-DHIS and HRIS-DHIS. The correlation between the number of HIV-positive clients and nurses and clinical officers in 2013 was $R^2 = 0.251$ and $R^2 = 0.261$ respectively. Wrong MFL codes, data type mis-match and hyphens in legacy data were key causes of data transmission errors. Lack of information exchange standards for aggregate data made programming time-consuming.

Keywords:

Information systems; Health information exchange; Interoperability; Health workforce; HIV.

Introduction

Sub-Saharan Africa (SSA) bears the heaviest burden of HIV. As of 2012, nearly two-thirds of the world's 34 million people infected with HIV lived in SSA [1]. Over the last ten years, there has been an unprecedented scale up of HIV prevention, care and treatment services. For example, UNAIDS reported a 40-fold increase in the number of HIV-infected persons receiving antiretroviral therapy (ART) from 2002 to 2012 [2]. A major obstacle to the scale-up of HIV care services in SSA is the chronic shortage of health workers [3, 4]. The health workforce density in the majority of countries in SSA fall below the World Health Organization (WHO) recommended minimum health worker per population [5].

In spite of support from multilateral and bilateral partnerships such as the Global Fund to fight AIDS, Tuberculosis and Malaria (GFATM) and US President's Emergency Plan for AIDS Relief (PEPFAR) as well as commitment from host-country governments, there is inadequate data to show that the number of deployed health workers, which include doctors, clinical officers, nurses and nurse-midwives, has increased with the scale up of HIV services. Studies from Zambia and Malawi show an increase in HIV-related workload against a relatively unchanged number of health workers [6, 7]. Data used in these studies were mainly health records and staff interviews from a few sampled health facilities. None of the studies used data from national health information systems such as the district health information systems (DHIS) or human resources information system (HRIS), and hence, had limited representativeness. A paper presented at the First Global Symposium on Health Systems Research, 2010 reported that low-income countries, which bear the heaviest disease burden, also have the weakest health workforce information systems [5].

Fully interoperable national health information systems are not yet common phenomena globally, and remain elusive in many countries due to a range of challenges including lack of standardization, lack of national unique identifiers, inadequate infrastructural capacity, lack of adequate skilled personnel, inadequate financial resources, and legal and organizational concerns [8-10]. Current initiatives focus on software development [11]. Published work also focuses on the software and on the exchange of health information at the individual patient level where interoperability has been shown to enhance continuity of care and efficiency [11, 12]. A few countries in SSA, including Kenya and South Africa, have developed national eHealth strategies prioritizing interoperable information systems as recommended by the WHO and the International Telecommunications Union [13]. However, the benefits of such systems is yet to be demonstrated. The need for high quality data from multiple interoperable sources for enhanced quality of care and to understand the correlations between scale-up of healthcare services and health workforce has never been greater, especially in resource-limited settings.

We conducted an observational study to assess the effect of interoperable national electronic health information systems on enhancing data availability to evaluate the effect of scale-up of HIV programs on human resources for health (HRH) in Kenya.

Methods

We created an environment of four interoperable systems that are part of the Kenyan eHealth architecture [14] as illustrated in Figure 1. The four systems were the district health information software-2 (DHIS2), regulatory human resource information system (rHRIS), integrated human resources information system (iHRIS), and the Master Facility List (MFL, a registry of health facilities in Kenya).

DHIS2: DHIS2 is a tool for the collection, validation, analysis and presentation of aggregate (not patient level) health statistics (https://hiskenya.org/). It is intended for, but not limited to, health information management activities. DHIS2 is a free and open source, web-based application. Kenya is among more than 30 countries in Africa, Asia, and Latin America, that have adopted DHIS2 as a part of their national health information system. Each month at the health facility level, aggregate health services statistics for all diseases, including HIV are entered into DHIS2. The automation of transmission of patient level data from electronic medical records (EMR) directly into DHIS2 is currently in progress. The aggregate data entered into DHIS2 include family planning, maternal child health, sexually trasmitted illneses, child health and nutrition, tuberculosis, and HIV services including HIV testing, prevention of mother to child transmission and antoretroviral therapy (ART). The data are stored in a central DHIS2 database hosted on a server at the Kenyan Ministry of Health (MOH) headquarters in Nairobi. DHIS2 implements application programming interfaces (APIs) which enable it to exchange information with other systems that implement similar or compatible APIs.

rHRIS and iHRIS: Human resources information systems (HRIS) collect and manage routine, national level, multi-cadre data on the health workforce including supply (i.e. training, exam, registration, licensure, intent to out-migrate, and continuing professional development) and deployment (i.e. health facility of deployment, date of appointment, work station in the facility, date of promotion, disciplinary actions, date of exit, and transfers). Regulatory human resources information system (rHRIS) collects and manages health workforce supply information while the integrated human resources information system (iHRIS) collects and manages deployment information. More information about the systems can be found at http://emorykenya.org/ and http://www.ihris.org/. A composite profile of a health worker can be created by linking the supply and deployment data detailing their educational, registration and employment credentials from both systems. rHRIS and iHRIS are web-based applications that allow updates to be made into secure databases by stakeholders in remote locations and for the generation of routine reports. The rHRIS and iHRIS have APIs that allow them to be interoperable with other systems such as DHIS2 and electronic medical record (EMR) systems.

MFL: WHO's guidelines for creating a master health facility list defines an MFL as a complete list of health facilities in a country, whether public or privately owned, and contains administrative information, identification information (signature domain) and service capacity (service domain) [15]. The set of identifiers in the signature domain uniquely identifies each health facility while information on service domain includes an inventory of services available and service capacity, which are essential for health systems planning and management. The MOH maintains an MFL that was created in 2008 by merging and reconciling several facility lists, which contained different, and sometimes conflicting, information about health facilities in the country. The MFL Code is a five digit number that uniquely identifies each health facility. Among the information contained in the MFL are ownership (Government of Kenya, private company or a faith-based organization), facility type (dispensary, health center, district hospital or referral hospital), administrative location (county, district, division, location), bed capacity, contact information (postal address and telephone number) and GIS coordinates (geo-coordinates). The MFL is updated regularly by the district health information and records officer whenever a new health facility is registered, an existing facility changes status, or if the information about it needs updating. As of November 2014, the MFL had 9,882 health facilities listed. The MFL database has an API that allows it to be interoperable with other systems such as the DHIS2, EMRs and HRIS.

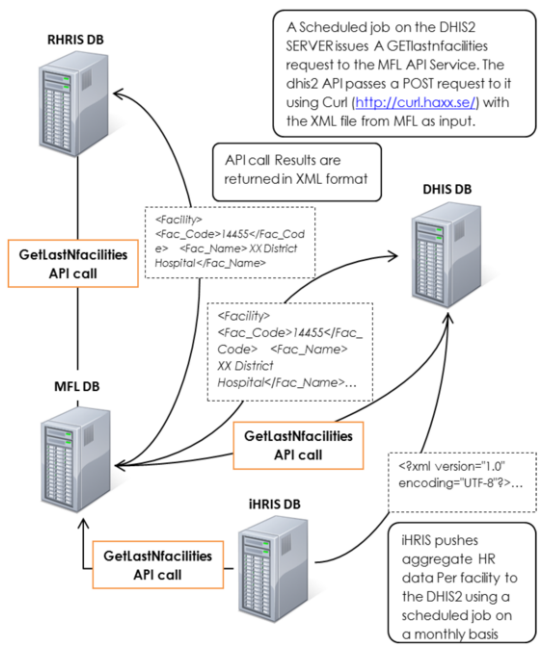


Figure 1: MFL-DHIS2; MFL-rHRIS; MFL-iHRIS; DHIS2-iHRIS Dynamic Interoperability

MFL-DHIS2-rHRIS-iHRIS Interoperability

The MFL API is currently implemented as a set of functions that accept defined parameters and generate details of facilities from the MFL in an eXtensible Markup Language (XML) that can be automatically parsed and read into a receiver database/system. More information on the exact implementation and usage of these functions is accessible at http://api.ehealth.or.ke. In order to make the MFL secure, a basic HTTP authentication has been implemented on the MFL API requiring a username and password to be passed into the MFL API for authentication purposes before any data is sent back to the receiver application.

In the MFL-DHIS2 implementation, an automated script that is scheduled to run daily at midnight using *cron* has been implemented to pull in the latest 100 facilities whose details are either new or have been added to the system. The ideal number to import from MFL depends on the nature of implementation and needs of the receiver system, however a random number of 100 was chosen initially and the basis of this selection was the fact that no more than 100 facilities have ever been added/modified in a day from across the country. This was determined by reviewing the facilities data based on date of addition and modification of records. It is however important to note that the higher the number of records selected, the slower the speed of import. To address the incompatibilites between legacy and current data, the legacy data were cleaned and data types modified to align with the programs and current data definitions.

iHRIS – DHIS2 integration: iHRIS-Manage system sends aggregate data on health workers per facility at the end of each month, disaggregated by county, cadre and gender, to DHIS2. This is done via scheduled *cron* jobs on the iHRIS-Manage server that automatically sends a preformatted file generated by the system using XLST (EXtensible Stylesheet Language).

An equivalent dataset was designed on DHIS2 with details of the data elements that hold the aggregated data values from iHRIS-Manage.

The rHRIS – DHIS2 integration is not yet fully automated. The DHIS team created a dataset for rHRIS to post aggregate training, registration, licensure and practice data. A data call to the rHRIS API pulls the selected data elements on health worker supply into the DHIS2 including the health-workers' registration status.

Error logs generated by scheduled batch jobs were used for tracking of exceptions that require attention.

Outcome Measures

To show the added value of interoperable systems in data availability and use, we created a flat file at the MOH's database populated with data abstracted from the interoperable data sources in order to understand the correlations between the national scale-up of HIV services and deployment of health workforce in Kenya. This data was hosted on the DHIS2 server and is accessible to key stakeholders in HRH, including the MOH and its partners and authorized users of DHIS2.

We used the following outcomes, (i) description of experiences implementing HRIS-MFL-DHIS2 interoperability (ii) correlation between scale-up of HIV care (measured through the number of HIV-positive individuals receiving HIV care) and deployment of health workers in Kenya, and (iii) trends in distribution of health worker per person living with HIV in Kenya. We also reviewed the data completeness as a component of data quality in the data sources.

Data Abstraction

After linking the MFL-DHIS2-rHRIS-iHRIS data, we abstracted the following data elements:

rHRIS/iHRIS	MFL	DHIS2
Cadre of health worker	MFL code	MFL code
MFL code	Facility name	Patients on care
Number of health workers	GIS coordinates	Patients on ART
Status of registration	County code	County code

From the above data elements, annual summaries by county were obtained, including the number of health workers deployed (nurses and clinical officers), number of patients enrolled on HIV care and number of patients currently receiving ART for 2012 and 2013. Additional data on the total number of people living with HIV per county were obtained from the national HIV estimates for Kenya based on mathematical modelling [16].

Due to low reporting rates in DHIS2 for the routine health statistics in 2011, we did not consider data for 2011 in the analysis.

Statistical and Spatial Data Analysis

We used scatter plots to assess the correlation between the number of patients receiving ART and the number of nurses in service. *Stata* was used to perform the analyses and calculate R^2 values. We used *ArcGIS*, a spatial analysis tool, for visual presentation of ratio of nurse to people living with HIV per county.

Ethical Considerations

We used aggregate data that did not contain any individually identifiable patient data. Additionally, the identifiable information of the health workers were removed by the MOH and regulatory boards/councils after linking the HRIS data to the DHIS2 and MFL. Dummy identifiers were assigned in order to conduct the analyses. There was no contact between the study team and patients or the health workers.

Results

Experiences implementing interoperable MFL-DHIS2-rHRIS-iHRIS

The synchronization mechanism described using the API has been largely successful; over 95% of data were successfully and accurately transmitted from the MFL to DHIS2 and from iHRIS to DHIS2. From the error logs, we were able to identify the most common causes of errors that prevented data from being imported into receiver systems (e.g. DHIS2). The common causes of errors include:

- Use of wrong MFL codes especially for older/legacy data that existed in the systems before automation of the data exchange process
- Data type mismatch that occurs when a system receives a request for a data type that is different from that of the stored variable (e.g. if a system received a request for a numeric variable when the stored variable is of type string)
- Presence of hyphen or an apostrophe in the data source.

The above sources of error were addressed through ongoing cleaning of legacy data and alignment of data types in the data files and programs.

Programming time for each system to ensure the APIs could exchange data was time consuming. It took nearly 10 months of programming, testing and documentation to achieve working solutions, and another month to get XML and XSLT transformations of facility list data between MFL and iHRIS right.

Data availability and use

Data was readily available in DHIS2 and was abstracted to support the analysis below:

(i) Nurses deployed vs. the number of HIV-positive persons: A total of 1,565,505 and 1,599,565 persons were HIV-positive in 2012 and 2013 respectively. Of these, 400,768 and 547,579 patients were receiving ART at the end of 2012 and 2013 respectively. The total number of nurses in service was 17,604 and 26,399 in 2012 and 2013 respectively. There was a weak, but positive, correlation between the number of the number of nurses deployed and the number of HIV-positive persons, R^2 = 0.281 (2012) and R^2 = 0.251 (2013). Figure 2 below shows the scatter plots for the periods under consideration. Each circle represents a county and the size is proportional to the number of HIV-positive persons.

(ii) Clinical officers deployed vs. HIV-positive clients: A total of 3,209 and 3,284 clinical officers were in service in 2012 and 2013 respectively. The correlation between the number of patients receiving ART and the number of clinical officers in service was positive but weak: R^2 = 0.380 (2012) and R^2 = 0.261 (2013).

(iii) Nursing workforce density (ratio of nurses to people living with HIV (PLHIV): The ratio of nurses to PLHIV improved from 1:89 to 1:61 from 2012 to 2013. The map in Figure 3 shows the change in the number of nurses per PLHIV by county.

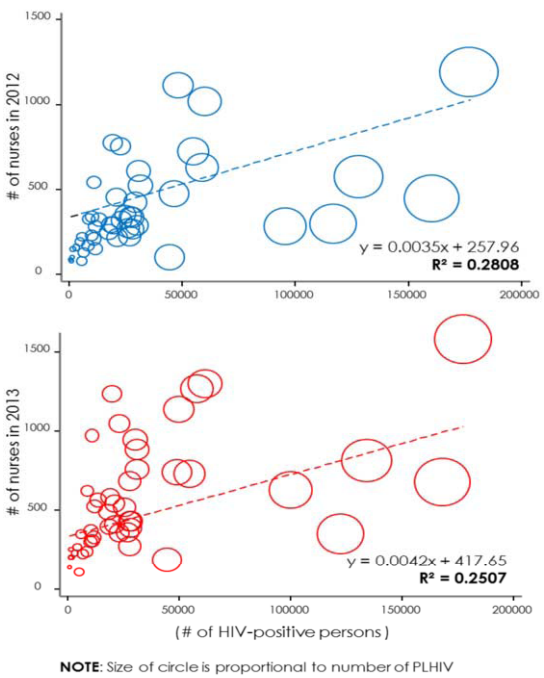


Figure 2: Correlation between number of HIV+ persons and no. of nurses by county in 2012 and 2013 in Kenya

Data Quality

Data quality varied based on the original source of data. Reporting rates for routine health service statistical summaries in DHIS2 improved tremendously from 2011 to 2012. In 2011, reporting rates to DHIS2 in a few counties were as low as 40% of the health facilities while in 2012, the majority of the counties had reporting rates above 80% of health facilities. Although coverage of the MFL data was high (>80% of health facilities in Kenya were listed in the MFL in 2012), a few facilities were missing from the MFL. There were no data losses during the transmission of data from the MFL to DHIS2 or the MFL to rHRIS/iHRIS.

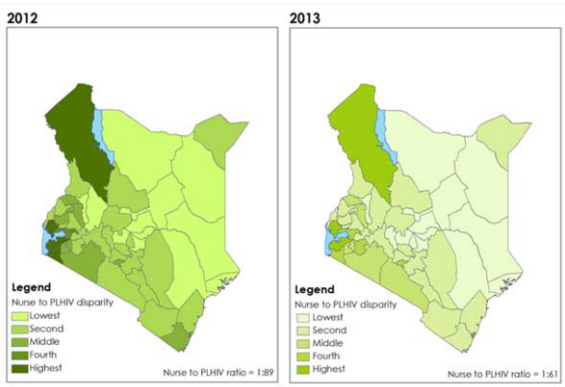


Figure 3: The ratio of nurse to person living with HIV by county in 2012 and 2013 in Kenya

Discussion

The implementation of interoperable MFL-DHIS2-iHRIS-rHRIS was successful. There was a weak, but positive, correlation between the number of people living with HIV and number of health workers in service, including nurses and clinical officers. Data quality varied by data source; the completeness of DHIS2 data was low in 2011 but improved in subsequent years. Over ninety-five percent of the data was transmitted from source systems (MFL and iHRIS) to the receiver system (DHIS2) making data readily available for analysis. The programming time and costs expended to achieve information exchange between the systems was high due to the lack of implementation of common standards and protocols that could be interpreted by the different systems.

The most common obstacles to the seamless synchronization of data between the different systems were identified as inconsistencies between legacy data and the current codes and data types. These findings are consistent with those reported by Alkaldi *et al* [17].To address these problems, we ensured that data cleaning and standardization of all legacy data was achieved before the automation of the data exchange using the API was done. In addition, appropriate validation rules or mechanisms were put in place to ensure that data from the source system is consistent with the type on the receiver system. The system administrators routinely reviewed the log files on the exchange process to ensure that data was exchanged as expected between the systems and took appropriate actions whenever errors were encountered. Although not implemented in this study, the use of data exchange schema and standards that enable *foundational, structural* and *semantic* interoperability defined by the Healthcare Information and Management Systems Society (HIMSS), such as XDS or the ADX data exchange protocol under development by the Integrated Healthcare Enterprise (IHE) - http://wiki.ihe.net/index.php?title=Quality,_Research_and_Public_Health, could further address the challenges encountered.

From the unweighted analysis, we found a weak but positive correlation between the number of patients receiving ART and the number of health workers (nurses and clinical officers) in service. This is similar to findings in Malawi and Zambia

which face similar human resource challenges as Kenya and other SSA countries [6, 7]. There was a marginal reduction in the ratio of health workers to the number of PLHIV, but with large variations geographically. Although we did not analyze the causes, mal-distribution of health workers could be informed by several factors including disease burden (HIV prevalence), urban/rural location and inadequate use of data to inform health workforce planning [18].

Our study had some limitations. We focused on the software component of interoperable systems, which is just one of the many factors that contribute to comprehensive information exchange of the national eHealth system. In order to implement these findings at scale, there is the need to tackle wider organizational, policy and infrastructural factors that affect the scale-up of interoperable health information systems [9, 10]. Low data quality, mainly due to low reporting rates by counties to the DHIS2, was identified as a key obstacle to data use. Although not presented with the results of this study, counties with fewer health workers were more likely to have incomplete data. It is worth noting that data completeness has significantly improved over time and currently reporting rates stand at over 80%. Finally, the data we presented in this study were not weighted or adjusted for factors that confound health workforce distribution such as population, disease burden and rural/urban locations.

The succcessful implementation of interoperable systems however provides an excellent opportunity for integrating data sources and enabling comprehensive analyses including health workforce and other health services data.

Conclusion

We demonstrated a successful implementation of interoperable MFL-DHIS2-iHRIS-rHRIS and showed added value in data availability and data use. There was a weak, but positive, correlation between the number of patients receiving ART and the number of health workers. More work needs to be done to assess the effect of use of information exchange standards on efficient achievement of interoperable systems as well as non-software factors associated with scaling up interoperable systems at a national level in resource-limited settings. Additionally, well designed studies are needed to understand correlations between the scale-up of HIV services and the health workforce in resource-limited settings.

References

[1] UNAIDS. Global report: UNAIDS report on the global AIDS epidemic 2013. 2013 Jan 11.
[2] UNAIDS. UNAIDS 2013: AIDS by the numbers. 2013.
[3] Egger M, Boulle A, Schechter M, and Miotti P. Antiretroviral therapy in resource-poor settings: scaling up inequalities? Int J Epidemiol 2005 Jun: 34(3): 509-12.
[4] Rasschaert F, Philips M, Van LL, Assefa Y, Schouten E, and Van DW. Tackling health workforce shortages during antiretroviral treatment scale-up--experiences from Ethiopia and Malawi. J Acquir Immune Defic Syndr 2011: 57(2 Suppl): S109-S12.
[5] Jimba M, Cometto G, Yamamoto T, Shiao L, Huicho L, and Sheikh M. Health workforce: the critical pathway to universal health coverage. 2010.
[6] Brugha R, Kadzandira J, Simbaya J, Dicker P, Mwapasa V, and Walsh A. Health workforce responses to global health initiatives funding: a comparison of Malawi and Zambia. Hum Resour Health 2010: 8: 19.
[7] Walsh A, Ndubani P, Simbaya J, Dicker P, and Brugha R. Task sharing in Zambia: HIV service scale-up compounds the human resource crisis. BMC Health Serv Res 2010: 10: 272.
[8] Kierkegaard P. Interoperability after deployment: persistent challenges and regional strategies in Denmark. Int J Qual Health Care 2015: Feb 19.
[9] Ovies-Bernal DP, and Gudelo-Londono SM. [Lessons learned in the implementation of interoperable National Health Information Systems: a systematic review]. Rev Panam Salud Publica 2014 May: 35(5-6): 415-23.
[10] Vest JR. Health information exchange: national and international approaches. Adv Health Care Manag 2012: 12: 3-24.
[11] Magrabi F, Aarts J, Nohr C, Baker M, Harrison S, Pelayo S, et al. A comparative review of patient safety initiatives for national health information technology. Int J Med Inform 2013: 82(5): e139-e48.
[12] Dogac A, Yuksel M, Erturkmen GL, Kabak Y, Namli T, Yildiz MH, et al. Healthcare information technology infrastructures in Turkey. Yearb Med Inform 2014: 9(1): 228-34.
[13] World Health Organizations, International Telecommunications Union. National eHealth Strategy Toolkit Overview. 2012.
[14] Kenyan Ministry of Health. Kenya National eHealth Strategy: 2011 - 2017. 2011.
[15] World Health Organizations. Creating a Master Facility List. 2013.
[16] Kenya National AIDS Control Council, UNAIDS. Kenya AIDS Progress Report. 2014.
[17] Alkhaldi B, Sahama T, Huxley C, and Gajanayake R. Barriers to implementing eHealth: a multi-dimensional perspective. Stud Health Technol Inform 2014: 205: 875-9.
[18] Buchan J, Couper ID, Tangcharoensathien V, Thepannya K, Jaskiewicz W, Perfilieva G, et al. Early implementation of WHO recommendations for the retention of health workers in remote and rural areas. Bull World Health Organ 2013: 91(11): 834-40.

Acknowledgments

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention or the Government of Kenya.

Address for correspondence

Tom Oluoch, US Centers for Disease Control and Prevention, Division of Global HIV/AIDS, Kenya. E-mail: toluoch@cdc.gov; oluoch.tom@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-682

Improving Clinical Decisions on T2DM Patients Integrating Clinical, Administrative and Environmental Data

Daniele Segagni[a], Lucia Sacchi[b], Arianna Dagliati[b], Valentina Tibollo[a], Paola Leporati[c], Pasqale De Cata[c], Luca Chiovato[c], Riccardo Bellazzi[b]

[a] *Laboratory of System Engineer for Clinical Research, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy*
[b] *Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy*
[c] *Unit of Internal Medicine and Endocrinology, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy*

Abstract

This work describes an integrated informatics system developed to collect and display clinically relevant data that can inform physicians and researchers about Type 2 Diabetes Mellitus (T2DM) patient clinical pathways and therapy adherence. The software we developed takes data coming from the electronic medical record (EMR) of the IRCCS Fondazione Maugeri (FSM) hospital of Pavia, Italy, and combines the data with administrative, pharmacy drugs (purchased from the local healthcare agency (ASL) of the Pavia area), and open environmental data of the same region. By using different use cases, we explain the importance of gathering and displaying the data types through a single informatics tool: the use of the tool as a calculator of risk factors and indicators to improve current detection of T2DM, a generator of clinical pathways and patients' behaviors from the point of view of the hospital care management, and a decision support tool for follow-up visits. The results of the performed data analysis report how the use of the dashboard displays meaningful clinical decisions in treating complex chronic diseases and might improve health outcomes.

Keywords:
Integrated Systems; Therapy Adherence; Clinical Pathways; Data Integration; Type 2 Diabetes.

Introduction

Treatment and management of chronic diseases often takes place outside clinical settings and impacts the daily life of patients. As a consequence, clinicians depend on patient reports of symptoms, side effects, functional status and treatment adherence. Patients typically report at clinical visits that are months apart, and recall accuracy can be highly fluctuating [1]. The possibility to gather and analyze accurate information at the right time from different sources, such as public healthcare systems, open data repositories, or hospital information systems is necessary to be able to perform correct clinical decisions [2]. The work presented in this paper aims to provide clinicians and researchers a software architecture based on data coming from heterogeneous data sources to support medical discovery and evidence-based practice about managing and controlling the evolution of chronic diseases. The system architecture relies on the open source i2b2 clinical data warehouse (CDW) [3] and the work done has been focused on expanding the basic functionalities of the framework by enhancing its visualization layer. A number of modules have been designed and developed to present the results through an intuitive and easy to use dashboard that shows results of process mining, temporal abstractions and similarity algorithms.

Materials and Methods

Type 2 Diabetes Mellitus (T2DM) is the most common form of diabetes. It accounts for at least 90% of all cases of diabetes. The World Health Organization (WHO) estimates that by 2030 there will be about 550 million people suffering from this disease [4]. This disease can remain undetected for many years because hyperglycemia (consequence of the insulin defects) develops gradually, and at earlier stages, the disease is not severe enough for the patient to notice any of the classic symptoms of diabetes. Despite the large number of models being developed and the increased interest and acknowledgement in the clinical field, only a small part of these models ends up being used in clinical practice [5]. One of the most important challenges of the MOSAIC European Union project is to combine the research activities related to the discovery of new risk factors, methods and models for diabetes onset, progression and evolution, with the development of software tools. The components and modules would incorporate innovations and make them usable by different end-users in a variety of settings. The biggest challenge of the project consists in translating innovations to achieve impact in the current clinical practice.

Data sources

To fill the gap resulting from infrequent clinical follow-ups of diabetic patients, the EMR used in FSM, a private hospital in Pavia, Italy, has been enhanced with data coming from the local public healthcare agency (ASL) of the Pavia area. This data contains administrative findings (e.g., prescription-based drug purchases), and data reporting environmental information (e.g., air temperature, air pollution, etc.) provided by the "Regione Lombardia" databases as open data.

Clinical usage

The main component of the software tool developed consists in a single page dashboard, where users can interact with a graphical presentation of clinical parameters. Dashboard users can analyze data though three different use cases:

- Use Case 1 (UC1): risk factors and indicators to improve current detection of T2DM

- Use Case 2 (UC2): hospital care management
- Use Case 3 (UC3): clinical decision support during follow-up visits

In this paper we will focus on use case 2 and 3, which are the ones that will be implemented for the management of already diagnosed T2DM patients at FSM. UC1 case is active in another research center participating in the MOSAIC project: the Hospital Universitari i Politècnic La Fe located in Valencia, Spain. UC1 will be integrated also in FSM by the end of the project, to allow a full evaluation of the system.

Architecture

We designed and developed a dashboard (Figure 1) to accomplish the goals described in UC2 and UC3. The dashboard communicates with the i2b2 DW where clinical, administrative and environmental data are integrated. A data mining module is responsible for running advanced temporal data mining algoritms on data queried from the CDW.

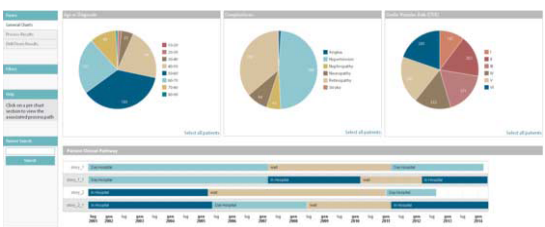


Figure 1 – starting page of the MOSAIC dashboard: charts show patients grouped by age at diagnosis, types of complications and cardio-vascular risk. The timeline displays four clinical pathways related to the selected patient set.

From this section, the user can either choose to select a subset of the population to use for further analyses or to search for an individual patient to examine his/her clinical charts. In the following paragraphs we will present some of the functionalities related to the two mentioned use cases.

Use Case 2.

UC2 is focused on the analysis of the clinical histories of a population of patients. This population could either be the whole patients' sample included in the data repository or a subset of patients conveniently selected by the user according to some clinical criteria available in the starting page of the dashboard. The models developed for data analysis rely on temporal and process mining techniques and are aimed at detecting the most frequent clinical patterns that are experienced by the patients' population [6, 7]. These models are focused on taking explicitly into account the temporal dimension, which strongly characterizes the evolution of chronic diseases. The ultimate goal of applying these techniques is to extract meaningful patients' stratifications that are based on the temporal evolution of some interesting clinical variables. For example, the most critical pathways in terms of severity that arise from complications and use of hospital resources could be highlighted. Temporal data mining is employed to understand which are the most interesting variables to use and to exploit the novel perspectives offered by temporal data integration.

Clinical histories are mined to extract behavioral patterns of prescription-related drug purchases, frequent clinical temporal patterns, cardio-vascular risk (CVR) profiles, level of complexity and evolution of complications. The extracted temporal patterns are presented to the user as timelines. The user can select a specific temporal pattern to drill down to the patients' population experiencing that behavior.

The target population of UC2 is of patients already diagnosed with T2DM. The data used to build and train UC2 models come from the FSM hospital EMR and are related to a cohort of T2DM patients followed up from 5 to 10 years. The EMR information related to these patients has been enhanced with public data coming from the Pavia ASL and with environmental information. This kind of data contains the history of all kind of therapies, hospitalizations, outpatient services, patient geo-localization, air temperature and degree of air pollution. To assess the complexity of care of the patients over time, a set of complexity stages, determined by the number of complications and hospitalizations patients undergo during the course of their disease, has been defined. The computation of these complexity stages is made possible by the availability of both clinical data related to the onset of complications and administrative information on hospitalizations and outpatient visits. Assuming that each patient starts with a stable state (it refers to pre-diabetic patients and type 2 diabetes patients who do not suffer any complications yet), when a complication arises the state shifts to “level 1.” If a second complication appears later than one year or if the second complication is a myocardial infarction, the state shifts to “level 2,” while if a second complication appears earlier than one year, the patient’s state shifts directly from “stable” to “level 2.” When an hospitalization, either due to the status of their diabetes-related complications or because of their metabolic instability occurs, the state change from “level 1” or “level 2” to "level 3.”

Use Case 3.

This use case refers to supporting clinicians to better manage T2DM chronic patients. Starting from patient’s EMR record, the system is able to project in a graphic way clinical information and behavioral patterns of prescription-related drug purchases. In this use case, clinicians are interested in having a better picture of what had happened to a single patient and what needs to be improved (e.g., changes in therapy, prescriptions of special examinations). Integrating information coming from the ASL and from the FSM EMR, analyzing and making them available in a unique tool, is a first step to achieve this goal. The part of the dashboard related to UC3 depicts data related to an individual patient, with a specific focus in the integration of data coming from different sources. Data integration is very important, as the clinical information related to the period that goes from the onset to the first follow-up at FSM is missing. The longer this period lasts, the more information becomes unobservable. This can be a limitation, especially if the objective of the analysis is to build predictive models (e.g. for complications prediction), as the events that occur in this not known time window might be the early determinant of a complication to arise. The data integration performed helps to mitigate this problem. In fact, it offers some additional information which comes from the public healthcare data. With this data, the information related to the not known time window becomes partially observable, given that the patient has been diagnosed after 2003.

The dashboard functionalities related to UC3 involve clinical and administrative data representation. Regarding clinical information, the dashboard reports temporal data related to (1) HbA1c values from laboratory exams performed within FSM hospital, (2) the evolution of the patient’s level of complexity and (3) the time course of the cardiovascular risk (Figure 2).

Figure 2 – MOSAIC dashboard example representing clinical valuesevolution over time.

One of the largest parts of information contained in the ASL data warehouse is related to prescription-based drug purchases by citizens at the pharmacy. Since this information is not directly available to the physician during clinical practice, it is very important to visualize it in a meaningful way in the dashboard. In particular, we have focused on drug purchase representation and on evaluation of the adherence to treatment behavior of the patient. Each drug purchase has been inserted in the i2b2 CDW using its Defined Daily Dose (DDD) that allows the expected number of therapy days related to that purchase to be computed. As the information about the actual drug dosages a patient should take is not available in our datasets, we use drug purchase as a proxy for estimating drug intake. Under this assumption, the DDD information represents the days supply for a specific purchase and we use DDDs to evaluate patients' behavioral patterns related to drug treatments. For each patient we retrieved the first prescription of a specific drug and divided the whole prescription period in semesters. For each semester we compute the sum of DDDs, which is an estimate of the total drug intake of the patient in that semester. We chose to use six months as an observation chunk, as in the current practice at FSM the average time between two consecutive follow-ups for chronic patients is one semester. Setting this temporal constraint gives the possibility of defining the granularity to observe modifications, adherence and continuity in drug treatments. As result a profile for each patient was constructed including the date of each purchase and the time period covered by the dispensed drugs, so to define the purchase pattern over time. In addition to the temporal profiles describing drug purchases, specific attention has been devoted to investigating the behavior of the patient in terms of treatment adherence. In recent years, several works studied the problem of harmonizing the integration of different healthcare information coming from heterogeneous sources while processing data coming from EHR or administrative databases [8]. One of the most common measures is the medication possession ratio (MPR) that indicates the percentage of received days supply divided by a period of time [9]. In our analyses, we have computed adherence to treatment using a validated index that corresponds to the MPR index for drug acquisition. The index is the Continuous, Single Interval Measure of Medication Acquisition (CSA). CSA index is calculated as the ratio: sum of DDDs over an observation period/length of the observation period (in our case one semester, 182 days). The CSA index indicates the percentage of days covered by the prescription. Besides DDDs, the information on CSA has also been included in the CDW. According to the CSA value, patients have been grouped into four clusters: no adherence (range of adherence between 0% and 40%), poor adherence (range between 40% and 80%), adherence (range between 80% and 100%), and over adherence (more than 100%). Patients belonging to the last group purchase, in a specific time window, more drugs than their prescription. The graphical representation of DDD values and therapy adherence percentage are shown in Figure 3.

Figure 3– MOSAIC dashboard example representing therapies DDD and adherence

Results

In this section we will present the results we obtained on a first sample of 433 patients using the methodologies described in the dashboard. Retrieving data from FSM EMR we had to deal with the fact that diabetic patients do not start to be followed at the hospital immediately after diabetes diagnosis. In Italy, this situation happens because this type of chronic patient is initially managed by the general practitioner (GP) who manages the patient until the GP believes it is more suitable for the patient to start being followed at the hospital outpatient service. Despite the wide variability of the period of time elapsed between onset of the first encounter, the majority of the patients start to be followed by FSM at a stable stage (73%). Analyzing the pattern of development of the first complication for our patients' sample and computing the time between the detection of the first complication and the first encounter at FSM, we observed how the diagnosis of the first complication occurs close to the first visit at the hospital. This is reasonable, if we consider that it is often the worsening of a patient's conditions that triggers a GP's decision to send a patient to the hospital outpatient service. As soon as the patient is managed by the hospital, the process that allows diagnosing possible complications is started, resulting in the recording of the complication in the EMR. Table 1 shows the distribution of the first complication for the patients who developed it after the first encounter at FSM.

Table 1 – Distribution of patients' first complication developed after first visit at FSM

Type of complication	% of development
Occlusion and stenosis of carotid artery	32.5 %
Fatty liver disease	25.8%
Nephropathy	12.8%
Retinopathy	9.7%
Neuropathy	8.2%
Peripheral vascular	6.9%
Chronic ischemic heart disease	4.1%

The administrative data related to drug purchases of about 433 patients considered in the analysis process consist in 244.190 records and a number of 21.055 distinct ATC codes. Classification of the purchased drugs on the basis of clinical relevance was the first type of analysis that we performed. In this way it was possible to reduce the variety in the data set to build simpler and meaningful scenarios based on treatment behaviors. To this end, the following ATC classes have been selected: A10 (drugs used in diabetes), B01 (antithrombotic agents), C02 (anti-hypertensives), C03 (diuretics), C10 (lipid modifying agents). The resulting codes have been further

grouped in 17 classes defined on the basis of medical knowledge and of the ATC-WHO system [10]. Table 2 details the number of patients and the number of box purchases per drug. Drugs that are not specific for the treatment of diabetes have been considered at a higher level of the ATC taxonomy, whereas drugs used to treat diabetes have been analyzed at a deeper level.

Table 2 – Knowledge-based classification of the ATC codes considered in our analysis (drugs not directly related to diabetes treatment)*

Drug	Patients	N° of Purchases
Alpha glucosidase inhibitors	85	331
Antihypertensive *	73	554
Antithrombotic *	338	3598
Phenformin_Sulfonamides	7	17
Diuretics *	157	1542
Gliptin	80	220
Incretin	23	79
Insulin	103	1381
LipidLowering *	303	3484
Metformin	352	3318
Metformin_Incretin	20	66
Metformin_Sulfonamides	73	709
Metformin_Thiazolidinediones	16	56
Repaglinide	98	638
Sulfonamides	150	1108
Sulfonamides_Thiazolidinediones	1	4
Thiazolidinediones	36	137

After computing the CSA values, we fixed a threshold (80%) to define if in a period, the patient purchased the appropriate dosage of the treatment, according to our assumptions. If the CSA is lower than 80% we associate a label to the semester indicating that there is Adherence = NO, otherwise we state Adherence = YES. We also represent the semesters during which there are no prescription as Adherence = INTERRUPTION and semesters where the purchases exceed the number of days in the time windows as Adherence = OVER. Table 3 summarizes the count of semesters on the basis of the CSA value.

Table 3 - Overall number of semesters classified as CSA >=80% (YES), CSA<80% (NO), CSA>100% (OVER), no prescription (INTERRUPTION)

Drug	YES	OVER	NO	INTERRUPTION
Alpha glucosidase inhibitors	39	34	253	5
Antihypertensive *	61	133	296	28
Antithrombotic *	875	555	1515	291
Biguanides_Sulfonamides	7	7	3	
Diuretics *	126	147	883	151
Gliptin			74	5
Incretin	58	66	92	6
Insulin	109	286	429	58
LipidLowering *	582	714	1587	228
Metformin	490	694	1932	202
Metformin_Incretin	21	23	21	1
Metformin_Sulfonamides	108	359	188	54
Metformin_Thiazolidinediones	18	10	28	1
Repaglinide	61	161	375	41
Sulfonamides	158	382	488	74
Sulfonamides_Thiazolidinediones	1	1	1	
Thiazolidinediones	32	23	80	2

Discussion

Despite the different values of CSA medians for different ATC groups, we analyzed the results in order to extract reliable and accurate patient's characterization on the basis of the general purchase behaviors. Our analysis has been focused on detecting main patterns of drug purchases. We identified the following subject classes:

- patients who purchase less drug boxes than the population taking the same drug class (CSAPatient<CSAPopulation)
- patients who purchase more drug boxes than the population taking the same drug and for whom the median CSA values stay below the 100%, not exceeding the recommended dosage (CSAPopulation <CSAPatient<100%)
- patients who purchase more drug boxes than the population taking the same drug but whose median CSA values stay above the 100%, exceeding the recommended dosage (100%<CSAPatient< CSAPopulation)

By identifying these groups it is possible to detect those patients who behave differently from the population of patients treated with the same drugs. A different purchase pattern could identify critical patients, independently from the CSA value. The developed procedure addresses the stratification challenge thanks to three consecutive steps, which are devoted to transform raw data into meaningful features and then using these features to in-depth tailoring specific patients' traits.

Step 1: Detection of subjects with statistically significant different purchase patterns from the population.

For each patient and each drug calculated, the median values of adherence for the whole prescription period were compared to the median value of the patient with the median value of the population has been performed (Wilcoxon Rank-Sum Test to asses is significant different, $p < 0.05$). If $p<0.05$, the patient was been assigned to one of two groups

Step2: Profiling of the $CSA_{Patient}$>$CSA_{Population}$ group.

For each patient belonging to this group and each drug check, if median values are <100% then classify as $CSA_{Population}$ <$CSA_{Patient}$<100% (ADHERENT) else if median values are >100% then classify as 100%<$CSA_{Patient}$< $CSA_{Population}$ (OVER ADHERENT)

Step3: Detect an overall purchase behavior

Each patient has the count of the number of drugs where the patient is UNDER - ADHERENT – OVER. If the behavior is detected for >50% over the total drug purchased then classify as: Under Adherent (LESS than POPULATION), Adherent (MORE than the population but less than 100%), Over Adherent (MORE than the population and more than 100%).

After we applied the process and retrieved the adherence stratification, we compared laboratory tests values and life style data in the different mined groups to understand if behavioral patterns of prescription-based drug purchases can be used as

biomarkers of specific clinical conditions of the patients who verify those patterns. The clinical variables (BMI, cholesterol and HbA1c), show significant differences ($p<0.01$, calculated with Kruskal-Wallis test) in the three groups. In particular, the patients identified as Adherent have better clinical values that the other groups. For BMI and Cholesterol, this difference becomes even more apparent ($p <0.01$, calculated with Wilcoxon test) when the Adherent group is highlighted and compared with the two Under or Over adherent groups considered together. This comparison emphasizes the clinical conviction that patients that not have a good adherence to therapy might suffer from metabolic disorders. Furthermore, considering other observations about patients diet, it is possible to detect better habits in adherent patients. During the whole observation period, adherent subjects have 118 observations of good diet habits, while they show bad habits only in 44 visits. The dashboard has been used to represent this information for UC3. Four data items are indicated: the median value of patient DDDs, the median value of group DDDs, the p-value resulted from Kruskal-Wallis test made to verify if there is difference between the patient and the group DDD median value, and an arrow used as visual indicator to display a positive or a negative difference between medians. Furthermore, we integrated from the ASL data warehouse, the cardiovascular risk index calculated as defined by the Italian Ministry of Health in the CUORE project [11] that takes into account values of gender, age, smoke habits, systolic pressure, cholesterol, HDL cholesterol, the diagnosis of diabetes and use of anti-of hypertensive drugs to produce the risk value.

Conclusion

In this paper, we presented an informatics approach that addresses the challenge of providing adequate clinical information such as therapy adherence, evolution of clinical pathways, and levels of complexity during inpatients encounters. The aim of our implementation is to improve the access to medical information through integrated informatics techniques necessary to gather data from different and heterogeneous sources such as hospital EMRs and public health or administrative records. By providing information in advance to physicians and not following the traditional approach based on questions to patients, we want to minimize the rise of bias and misunderstanding that can occur while patients report about medications and therapies they are following. With the introduction of the developed system into work practice and incorporating it into the hospital EMR, it will be easier to continually maintain the assessment of T2DM patients during their follow-up visits. Moreover, using the developed dashboard, clinicians are able to compare patient data they are visiting with a similar group on the basis of the variable they are interested. Analyzing data at the individual level and comparing them to similar group might have more impact in predicting how the clinical situation could evolve and might be useful for clinicians to perform correct clinical decisions.

Acknowledgement

This work is part of the MOSAIC EU project, funded by the 7th Framework Programme (http://www.mosaicproject.eu/).

References

[1] Kim C, Williamson DF, Herman WH, et al. Referral management and the care of patients with diabetes: the Translating Research Into Action for Diabetes (TRIAD) study. Am J Manag Care 2004;10(2 Pt 2):137

[2] Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science. 2010 Nov 5;330(6005):759-60.

[3] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010, 17(2):124-130.

[4] Centers for Disease Control and Prevention. Diabetes report card 2012. In: US Department of Health and Human Services, ed. Atlanta, GA: Centers for Disease Control and Prevention, 2012:1–4. http://www.cdc.gov/Diabetes/pubs/reportcard.htm. Last access: December 22, 2014.

[5] Kengne AP, Beulens JW, Peelen LM, Moons KG, van der Schouw YT, Schulze MB, Spijkerman AM, Griffin SJ, Grobbee DE, Palla L, Tormo MJ, Arriola L, Barengo NC, Barricarte A, Boeing H, Bonet C, Clavel-Chapelon F, Dartois L, Fagherazzi G, Franks PW, Huerta JM, Kaaks R, Key TJ, Khaw KT, Li K, Mühlenbruch K, Nilsson PM, Overvad K, Overvad TF, Palli D, Panico S, Quirós JR, Rolandsson O, Roswall N, Sacerdote C, Sánchez MJ, Slimani N, Tagliabue G, Tjønneland A, Tumino R, van der A DL, Forouhi NG, Sharp SJ, Langenberg C, Riboli E, Wareham NJ. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. Lancet Diabetes Endocrinol. 2014 Jan;2(1):19-29.

[6] Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. J Biomed Inform. 2014 Sep 16.

[7] Sacchi L, Dagliati A, Bellazzi R. Analyzing complex patients' temporal histories: new frontiers in temporal data mining. Methods Mol Biol. 2015;1246:89-105.

[8] Raebel MA, Schmittdiel J, Karter AJ, Konieczny JL, Steiner JF. Standardizingterminology and definitions of medication adherence and persistence in researchemploying electronic databases. Med Care. 2013 Aug;51(8 Suppl 3):S11-21.

[9] Kozma CM, Dickson M, Phillips AL, Meletiche DM. Medication possession ratio:implications of using fixed and variable observation periods in assessing adherence with disease-modifying drugs in patients with multiple sclerosis. Patient Prefer Adherence. 2013 Jun 12;7:509-16.

[10] The ATC/DDD Index 2015: http://www.whocc.no/atc_ddd_index. Last access: December 22, 2014.

[11] Donfrancesco C, Palmieri L, Cooney MT, Vanuzzo D, Panico S, Cesana G, FerrarioM, Pilotto L, Graham IM, Giampaoli S. Italian cardiovascular mortality charts of the CUORE project: are they comparable with the SCORE charts? Eur J CardiovascPrev Rehabil. 2010 Aug;17(4):403-9. 1995;2:316 –322

Address for correspondence

Daniele Segagni: daniele.segagni@fsm.it

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-687

Architecture to Summarize Patient-Level Data Across Borders and Countries

Luís A. Bastião Silva [a], Carlos Días[b], Johan van der Lei [c], José Luis Oliveira [a]

[a]DETI/IEETA, University of Aveiro, Portugal
[b]Synapse Research Management Partners, Barcelona, Spain
[c]Erasmus University Medical Center, Rotterdam, Netherlands

Abstract

Translational medicine is becoming fundamental in promoting information flow between basic research and clinical practice, and in improving patients' health. The need for efficient systems to process and share information from multiple sources – including distinct areas of medicine – is a pressing need in biomedical research. Nevertheless, healthcare information systems are fragmented over different databases, medical institutions, and geographical locations.

There are already several approaches to tackle this problem based on centralized or distributed solutions. Nonetheless, mainly due to privacy reasons, these models only work for specific silos. In this paper, we present a new ecosystem for connecting database owners and researchers. Our approach consists of gathering, via a common fingerprint, an extensive characterization of dispersed databases. This fingerprint typically contains high-level aggregated information addressing questions at a population level and allowing, for instance, quick identification of databases with data that may help to answer a specific research question.

This work is being conducted in the context of EMIF (European Medical Information Framework), an IMI (The Innovative Medicines Initiative) joint undertaking funded project, wherein the Catalogue is being used to collect data from cohorts and Electronic Health Record systems of several European countries.

Keywords:

Translational Research; Real-World Evidence; Observational Studies; Cohorts.

Introduction

The quantity of clinical information and disease-specific data has steadily increased over the last decade. This information is fragmented over dispersed databases in different clinical silos around the world. However, as awareness about the potential of these data for clinical research increases, there is a growing need for solutions for secure sharing of information across different databases. This re-use of data will lead to many benefits, mainly for clinical and pharmacological researchers [1-3].

Several partial solutions have been proposed to solve the problem, from centralized to federated approaches with distinct security levels and different access roles. However, clinical researchers struggle to find datasets or clinical trial candidate subjects that fulfil their specific research questions. Moreover, this problem is not only due to privacy issues, but because most of the time appropriate datasets are not easy to locate, or their very existence is unknown [4].

In this paper, we propose a system that is able to summarize and aggregate databases' content, which is not dependent on the database type. While actual patient-level data are not shared and do not flow outside the healthcare institution's boundaries, clinical researchers are able to query high-level information and characteristics of the databases.

The system allows creating fingerprint templates, such as a set of questions that should adequately characterise each database type (e.g. observational data sources). Furthermore, researchers are able to search in the catalogue in any field. The integration of fingerprints from different databases and silos allows researchers to find databases suitable to their needs from a single access point.

This catalogue is being developed in the EMIF project (European Medical Information Framework), the broader main goal of which is to develop an information framework capable of integrating patient health information from different healthcare and research communities at a level that is currently not available.

Translational Research

The need for efficient systems to process and summarize information from multiple sources, including distinct areas of medicine, is a pressing need in biomedical research [3]. Translational medicine is fundamental to promote the flow of information between basic and clinical scientists, improve new biotechnologies, and increase patient health [5]. Biomedical informatics encompasses a set of methodologies to support knowledge transfer and integration across the major areas of translational medicine, from molecules to populations [2].

There are several questions that translational medicine will help to answer, starting by discovering which interventions might have an adequate benefit-risk profile in a given individual. Moreover, there are still many practices and protocols that are applied in clinical care based on intuition. Translational medicine can help researchers understand the reasons why some methods work, and how they can be developed in order to improve quality of healthcare. These results can be achieved through the contribution of different areas, including biomedical informatics.

Many areas may contribute to translational medicine, such as bioinformatics, imaging informatics, clinical informatics, and public health informatics. Bioinformatics seeks to improve the understanding of biological processes and therefore allow more targeted and efficient clinical interventions. Imaging informatics plays a significant role in understanding pathogenesis and identification of treatments at the molecular, cellular, organ, and tissue levels. New methods for visualising and analysing these data will be needed and are currently being developed [6]. Clinical informatics innovations are necessary to improve patients' care through the availability and integration of relevant information. Several methods are applied to medical report information, such as text mining and knowledge extraction. Public health informatics is the area

that intends to analyse and study new techniques to exploit data on a large scale. Translational medicine teams need to address many challenges, and it benefits from innovations in decision support, natural language processing, standards, information retrieval, and electronic health records, among other research areas.

There are efforts to integrate health data from multiple institutions, correlate it with genomic information, and incorporate knowledge from the scientific literature. Nevertheless, the integration of biomedical data is still largely not applied in practice, and significant obstacles pertaining to harmonisation challenges, ethical, legal and social issues around data sharing, the inherent complexity of the data, and difficulty in identifying appropriate datasets still hamper research and its translation to patient benefit.

Related Work

Many tools have been created to analyse clinical and genomic data, mainly for specific purposes, which allow researchers to extract knowledge from clinical databases. Consequently, several efforts have been made to combine biomedical data from different sources, with emphasis on clinical data integration.

REDCap (Research Electronic Data Capture) [7] provides a scientific research workspace for translational research. It allows users to quickly create and manage online surveys and databases. Users can insert data related to a research study and the forms are only available to users for whom access rights are granted. The software supports exporting the data to SPSS, R, SAS and Stata. It is a web-based application that has the ability to audit and log users' actions in the platform, and it can be used upon request. REDCap is being used in the production of more than 30,000 research studies, with over 92,000 users.

Mini-Sentinel [8,9] is a project aiming to create an active surveillance to monitor the safety of FDA-regulated medical products. It uses pre-existing electronic healthcare data from multiple sources. Several institutions provide access to data, as well as scientific and organizational expertise. The tools developed provide techniques to identify and validate medical products' exposure and potentially associated health outcomes. There are several isolated software tools to support specific research questions, allowing users to do research mainly inside the institution. However, there is still a great need for expertise to set up all these technologies.

ENCePP is a collaborative scientific network, launched by the European Medicines Agency (EMA), with the aim of strengthening the post-authorization monitoring of medicinal products, whilst bringing together relevant research centres, healthcare databases, electronic registries and existing networks covering certain rare diseases, therapeutic fields, and adverse drug events of interest across Europe. This project was completed in 2009 and provides a framework for improved pharmacoepidemiological research and post-authorization safety surveillance of medicinal products in Europe. This can be achieved by directly contacting the individual centres and networks registered in ENCePP or by submitting a request to the ENCePP secretariat to place an announcement in the ENCePP partners' forum. In the first phase of the project, it was possible to identify a number of suitable centres across Europe, leading to the establishment of a general inventory of research institutions following a survey at the level of EU Member States. In 2010, the ENCePP Database of Research Resources, which is part of ENCePP, was launched. The Resources Database is an electronic index of available EU research organizations and data sources in the field of pharmacoepidemiology and pharmacovigilance. It is available through the ENCePP web portal to the general public, and the information contained in this database is entered and maintained by the research organizations and/or data providers.

Bridge-To-Data [10] offers a tool that allows users to identify key features and compare database profiles. This resource also serves as an educational tool for public health research and as a template for health system planners to design or refine their healthcare data. BRIDGE provides a unique, searchable, and comprehensive compendium of information on population healthcare databases worldwide, allowing the healthcare professional to obtain profiles on various population datasets on a single website. The database profiles are updated with the most current data and verified by the respective database managers or representatives. Updated database profiles replace older versions to avoid duplicate reporting. BRIDGE subscribers can perform robust and advanced searches for a wide range of database profiles using the search page. Standardized data from each database allows subscribers to compare in-depth details from multiple databases. Informative resources are provided, such as a listing of recent publications using the relevant database and contact information for database managers. As part of its user-friendliness, BRIDGE also has a glossary, including coding systems to guide new subscribers with their search.

While these are already examples of solutions for database profiling, they do not apply deep summarization and there is still a gap in the clinical research communities, which are difficult to connect. A system for collecting detailed summarized views and asking high level questions regarding the databases is still needed.

Methods

A key idea behind the creation of the Catalogue is to allow researchers to find specific databases aligned with their research purposes. The main conception of the Catalogue was to have a summarized overview of a number of geographically scattered healthcare databases, with the aim of facilitating initial assessment and selection for specific research questions. Thus, a system that allows database owners to fill out a questionnaire and create their own database fingerprint was developed. Nevertheless, a process will be previously required to create a template schema of the fingerprint, which contains several questions with relevant information about the database and does not disclose person-sensitive information. The template schema of the fingerprint needs to be defined for each type of database, such as Observational Databases, Research Cohorts, or others. Thus, the Catalogue also allows system administrators to dynamically modify the questionnaires of each data source type.

Based on the appropriate template, database owners are able to fill in their database fingerprint, as shown in Figure 1 (i.e. answers to the questionnaire corresponding to their database type):

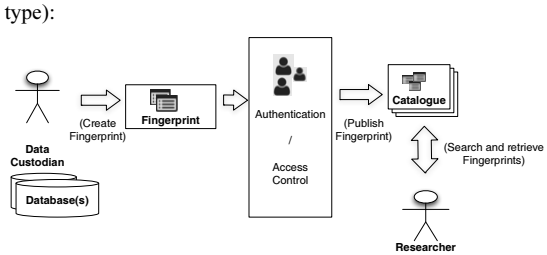


Figure 1: Workflow Description in the Catalogue

Fingerprinting a Database

While real data are not shared and do not flow out of healthcare institutions' boundaries, clinical researchers are able to query high-level information on the databases. We present the concept of the fingerprint, which is a questionnaire with a set of questions about a database type, such as database population, coding system, geographical distribution, and other relevant information.

Each template contains several groups of questions. Within these, each question can be individually defined, including the answer type (for instance, multiple-choice, plain-text, yes or no). Administrators can change each question or set of question and those entries are available in the administration panel (accessible for users with administration profile). Afterwards, the application will be automatically updated.

Users can create fingerprint templates using an online assisting tool, or a tabular textual schema (e.g. in Microsoft Excel). The process of building this kind of schema and allowing collaborative members to contribute without a steep learning curve was a priority when developing the Catalogue.

User Profiles

The definition of profiles and their role in the Catalogue is a really important issue. The user should have access to the features and interfaces of interest according to the research question, and therefore flexibility was a priority. We considered three types of users and defined their respective requirements: Researcher, Data custodian and System administrator.

The researcher is able to search over a set of fingerprints. A free-text search feature is provided, so that regular users can access all content provided by database owners. Comparison of search results, tabular views, and searching for 'similar' databases are also provided. This requires the definition and agreement of similarity measurements for specific property matches or various other properties, such as location proximity.

Data custodians have full control over their workspace. They are able to manage and access the databases they maintain on the system, and navigate through the corresponding fingerprint data. Due to the complexity of the database information, the user is also able to save the database fingerprint at any point and edit it over time.

The administrator is able to create and edit fingerprints schemas, create group questions in the templates, define questions, validate user registrations, and several other management tasks.

Architecture

The Catalogue is built in three distinct layers (Figure 2). The top layer is open for interactions with both end users, through the user interface, by third party applications, and through direct data exchanges with the Catalogue Web services. At this level, four main modules were developed: Browse Catalogue, Fingerprint Template, Administration Management, and API Web services.

The idea of the Browse Catalogue module was to create a workspace where the user is able to browse the fingerprints and also search for researchers' questions. Furthermore, the system allows the user to search in the Catalogue (search engine) and also compare results (e.g. compare two databases and return the similarity for each field in the fingerprint).

Fingerprint Templates is a module to create and edit the templates. Furthermore, an Excel template was defined to allow specific types of databases to easily define suitable new fingerprint templates so that they could be easily imported into the system. Also, data custodians can submit information about their database to the Catalogue. After the database fingerprint is submitted to the system, it should be searchable in the Catalogue.

Administration Management is available to administer the accounts and roles in the Catalogue. It also controls the user's access and registration in the system.

The API Web Services provide a set of programmatic endpoints that can be consulted by third party applications and by the Catalogue presentation layer. The main idea is that other applications can send data to the Catalogue in the format of key-value pairs, containing, for instance, population characteristics.

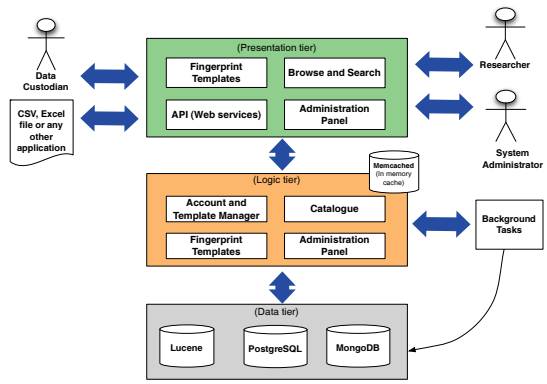


Figure 2: Catalogue - Software Architecture

There is also the logic layer containing the models that define the Catalogue's entities, such as users with different profiles and permissions, definition of the template schemas, fingerprint instances, and many others. The business logic is also defined in this layer. Finally, the data layer is where the information is stored. For example, the information regarding the catalogue templates and the dynamic groups of questions, as well as users' accounts, are stored in a PostgreSQL database. Furthermore, each database is indexed in an Apache Solr instance. This open source search engine framework allows full-text search and is used in large-scale information retrieval projects. Moreover, specific aggregated views of the database, such as population characteristics, are uploaded to a MongoDB instance, which allows greater flexibility in the stored values while maintaining good filtering features over stored data.

The Catalogue was developed in Python using Django[1], a web framework that encourages rapid development and clean programmatic design. However, a considerable part of the development was implemented in HTML5, CSS, and Javascript; namely, the interface and the end-user interaction. Furthermore, in order to improve the web design quality, we have adopted the Bootstrap[2] framework, a front-end aesthetic framework for web development.

Due to the high number of tasks the Catalogue is subject to, we realized that several processes could slow down the system when dealing with numerous concurrent interactions. Accordingly, we adopted several strategies to improve its performance. One of those is to have a cache based on the memcached[3] system. Several tasks, such as indexing the fingerprint in Solr, take too much time, making users wait for the opera-

[1] https://www.djangoproject.com
[2] http://getbootstrap.com
[3] www.memcached.org

tion to complete. All of these heavy tasks are added to a queue message (Rabbit MQ[4]) and the Celery backend executes them in the background.

Results and Discussion

EMIF (European Medical Information Framework) is a joint IMI project supported by the EU and the European pharmaceutical industry (through EPFIA), which has the overall ambition of enabling efficient re-use of health data for research purposes. It will allow the creation of new features to re-shape the way researchers answer key questions, and it will also determine possible future research directions [11]. The framework being developed (EMIF-Platform) supports two important research topics that serve as exemplars and test cases: 1) obesity and its metabolic complications (EMIF-Metabolic), and 2) markers for the development of Alzheimer's disease and other forms of dementia (EMIF-AD). Through the participating data sources, the project also aims to explore how the massive data available in pan-European Electronic Healthcare Record systems (EHR) and research cohorts can be optimally leveraged to improve biomedical research.

In this context, there is a need to integrate medical records, clinical and other omics information from different sources. Aiming to tackle this problem, a system that should be able to provide the first step towards answering questions from clinical researchers, in terms of knowing and comparing data sources' capacities, was created.

The current version of the EMIF Catalogue is accessible at http://bioinformatics.ua.pt/emif, but is only for registered users until a full governance model is set up in the project, which is expected to be completed by the end of 2017. Two distinct communities are already using the EMIF Catalogue:

- Observational Data sources: the fingerprint schema has 12 groups of questions, and in total 212 questions.
- Alzheimer Cohorts: the fingerprint schema has 27 groups of questions, with a total of 558 questions.

The EMIF Catalogue has already collected 46 fingerprints of different Alzheimer's Cohort databases and 15 from Electronic Healthcare records (EHR) data sources, corresponding to a total of around of 30 million patients. Overall, the Catalogue has more than 100 users registered, including both database owners and researchers.

Conclusion

Biomedical researchers have faced several difficulties in finding databases or information that fulfil their requirements.

In this paper, a software architecture able to fingerprint any kind of healthcare and research database was presented. The goal is to facilitate the initial assessment and selection of appropriate data sources for studies intending to answer a research question. The system is already being used to collect fingerprints from a wide collection of cohort and EHR databases, allowing faster identification of databases and resources.

Acknowledgments

This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant n° 115372). LBS is funded by FCT (Fundação para a Ciência e a Tecnologia) under grant agreement SFRH/BD/79389/ 2011.

References

[1] A. J. McMurry, S. N. Murphy, D. MacFadden, G. Weber, W. W. Simons, J. Orechia, *et al.*, "SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies", *PLOS ONE,* vol. 8, p. e 55811, 2013.

[2] I. N. Sarkar, "Biomedical informatics and translational medicine", *Journal of translational medicine,* vol. 8, p. 22, 2010.

[3] S. Szalma, V. Koka, T. Khasanova, and E. D. Perakslis, "Effective knowledge management in translational medicine", *Journal of translational medicine,* vol. 8, p. 68, 2010.

[4] M. Cases, L. I. Furlong, J. Albanell, R. B. Altman, R. Bellazzi, S. Boyer, *et al.*, "Improving data and knowledge management to better integrate health care and research", *Journal of internal medicine,* vol. 274, pp. 321-328, 2013.

[5] E. Bonsma and J. Vrijnsen, "Homogenising access to heterogeneous biomedical data sources", in *Biomedical Informatics and Intelligent Approaches in the Support of Genomic Medicine (BMIINT)*, 2009, pp. 12-23.

[6] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, *et al.*, "Visualization of omics data for systems biology", *Nature methods,* vol. 7, pp. S56-S68, 2010.

[7] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support", *Journal of Biomedical Informatics,* vol. 42, p. 377, 2009.

[8] R. Platt and R. Carnahan, "The US Food and Drug Administration's Mini Sentinel Program", *Pharmacoepidemiology and Drug Safety,* vol. 21, pp. 1-303, 2012.

[9] M. A. Robb, J. A. Racoosin, R. E. Sherman, T. P. Gross, R. Ball, M. E. Reichman, *et al.*, "The US Food and Drug Administration's Sentinel Initiative: expanding the horizons of medical product safety", *Pharmacoepidemiology and Drug Safety,* vol. 21, pp. 9-11, 2012.

[10] *B.R.I.D.G.E TO DATA*. http://www.bridgetodata.org/

[11] H. Zetterberg, "Unresolved questions in Alzheimer's research: will biomarkers help?", *Biomarkers in medicine,* vol. 8, pp. 61-63, 2014.

[4] www.rabbitmq.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-691

A Cloud-Based Infrastructure for Feedback-Driven Training and Image Recognition

Mani Abedini[a], Stefan von Cavallar[a], Rajib Chakravorty[a], Matthew Davis[a], Rahil Garnavi[a1]

[a] *IBM Research – Australia, Carlton, VIC, Australia*

Abstract

Advanced techniques in machine learning combined with scalable "cloud" computing infrastructure are driving the creation of new and innovative health diagnostic applications. We describe a service and application for performing image training and recognition, tailored to dermatology and melanoma identification. The system implements new machine learning approaches to provide a feedback-driven training loop. This training sequence enhances classification performance by incrementally retraining the classifier model from expert responses. To easily provide this application and associated web service to clinical practices, we also describe a scalable cloud infrastructure, deployable in public cloud infrastructure and private, on-premise systems.

Keywords:

Machine Learning; Information Processing; Information Storage and Retrieval; Cloud Computing; Dermoscopy; Melanoma.

Introduction

New and inexpensive computational and communication capabilities continue to drive growth in the amount of healthcare data generated. For example, the number of mammograms in the US has increased from 26 million in 2003 [1] to over 48 million today, resulting annually in many petabytes of new image data alone. With the growth of other heterogeneous types of data (text, images, video, genetic sequences), in a broad range of clinical domains, the need for new techniques and applications to store, process, and utilise these data continues to evolve.

Making sense of these large amounts of data is a complex but essential task. Algorithmic machine learning techniques are essential to extracting useful information or modelling large data sets [2-4]. In the context of clinical diagnosis – two major classes of machine learning algorithms exist: Fully automated algorithms for diagnosis and algorithms that assist an expert clinician to perform the diagnosis, or computer assisted diagnosis (CAD). The purpose of computer-based systems is to support clinicians in their decision making, not to replace them. Therefore, CADs have better synergy in clinical environment, by providing a suggestive diagnosis or "second opinion" to assist a clinician [5] and are already used in many routine clinical tasks [6,7].

Modern communication technologies permit access to computing resources (processing, data storage, etc.) remotely. This has significant benefit for clinical applications, making data collection, analysis, and reporting more flexible and reliable [8]. Application of machine learning algorithms on a large amount of data require massive computational infrastructure. "Cloud" computing provides a suitable model to access such computational systems in different levels – infrastructure, platform, software, and data. Despite some challenges, such as patient data privacy and associated regulatory concerns, cloud technologies are being adopted by organisations at an increasing rate [9,10]. For health care providers, it presents a cost-effective way to get access to advanced technologies [9]. Cloud computing platforms hide the complexity of multiple tools, operating platforms, and frameworks from the end user [11,12]. Moreover, care providers in remote regions can obtain access to specialized resources (human experts, databases, digital models etc.) in real time [13].

In the health domain, machine learning algorithms require large volumes of training data in order to draw useful conclusions. Designing these algorithms to run efficiently on cloud-provided infrastructure is an effective strategy. There have been efforts to develop frameworks, such as MapReduce [14] and GraphLab [15], which parallelize the execution of processing algorithms on the kinds of distributed infrastructure found in cloud environments. While these frameworks are more suitable for machine learning researchers than non-technical end-users, hosted software applications, built on the same compute infrastructure and tailored for clinical users, are becoming available [16,17].

In this paper, we introduce a new, feedback-driven application for performing image training and recognition, targeting dermoscopy image recognition and melanoma identification. This application provides a web application and web service to clinical practices, potentially scaling to thousands of users, while hiding the complexities of a modern machine learning system from medical domain experts.

Methods

The system is composed of two high-level components, a machine learning and analytics pipeline for performing training and recognition and a scalable, cloud-based infrastructure for storing models and media and for executing tasks against the analytics pipeline.

Image analytics and feature extraction

A "supervised" machine learning work flow (described in previous section) uses "features" or "properties" of images from various classes to build a model [18]. These features are collected from the image by applying a number of image processing algorithms.

1 Author order is alphabetical

The collected features can roughly be divided in three broad categories: Low, mid, and high-level features.

Low-level features are directly extracted from image pixel data. Histogram of pixel values or other empirical modelling of pixel intensity distribution is one such low level feature.

Mid-level features try to abstract some properties of the object of interest, collected from an image annotation. Typically, operations or processes take place on raw pixel intensities that result in higher-level features. One such example is "edge" [19], roughly a gradient operation on raw pixel intensities. Horizontal, vertical or diagonal edges or any combination of these edges can be easily obtained. Similarly, operators to detect corners, textures, and shapes are also available [20].

High-level features represent domain specific properties of a given image. For example, in a cardiac MRI image, an enlarged myocardium, or an infarct scar, can be a high-level feature.

Based on the expert opinion (e.g. annotation from domain experts), features are extracted either from an entire image, partial image, or both. This ensures that local and/or global features are captured from a single image.

The system allows the user to choose from a number of pre-defined features to be extracted from the image. The choice can be made available from the user interface or via a configuration file.

Analytics Pipeline

The analytics pipeline contains four subcomponents, as shown in Figure 1. Three of these stages are derived from steps in a traditional machine learning pipeline; namely, annotation, training, and classification. The pipeline also incorporates a fourth stage, which takes user feedback from a classification result to refine the classifier through incremental training.

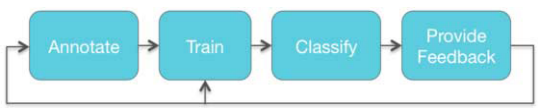


Figure 1 - Stages of the analytics pipeline

Annotate

For automated analysis and recognition of images of a certain class/type, annotation is an important step and prerequisite to training. In the annotation phase, a domain expert generates ground truth and training samples for the training module. In other words, annotation is the process of attaching metadata to images to facilitate development of classification models. In a computer-assisted system for melanoma recognition, for instance, the annotation can include:

- clinical diagnosis or histopathology of the given image (i.e. melanoma vs. benign)
- spatial boundary of the lesion, separating the lesion from surrounding normal skin
- number and types of colours (e.g. light brown, dark brown, black, etc.) observed in the lesion
- identified dermoscopy patterns (e.g. reticular pattern, globular pattern, pigment network, etc.) observed in the lesion

Clinical approaches, such as pattern analysis [21], ABCD rule of dermoscopy [22], CASH algorithm [23], and others, can be used as a basis for defining an "annotation template" and applied by the domain expert (clinician) in the annotation phase.

Our system provides an annotation image tool, as shown in Figure 2, to simplify the process of defining annotations for an image. The annotation tool enables a clinician to label features in an image from one or more templates provided through configuration. Labels are applied to the full image or spatial regions of interest, such as the lesion border. In addition to feature labelling, the clinician also specifies the clinical diagnosis for each image (melanoma or benign).

Train

After attaching annotations to training samples in the annotation phase, the clinician can train and build a classification model. First, the user creates a new training dataset by selecting from one or more pre-annotated images. A training set will contain images of mixed diagnosis, for example images with benign lesions and those diagnosed as melanoma. After creating the training set, the user launches the training algorithms by selecting "Train New Model". While the system defaults to a pre-configured set of parameters, advanced users

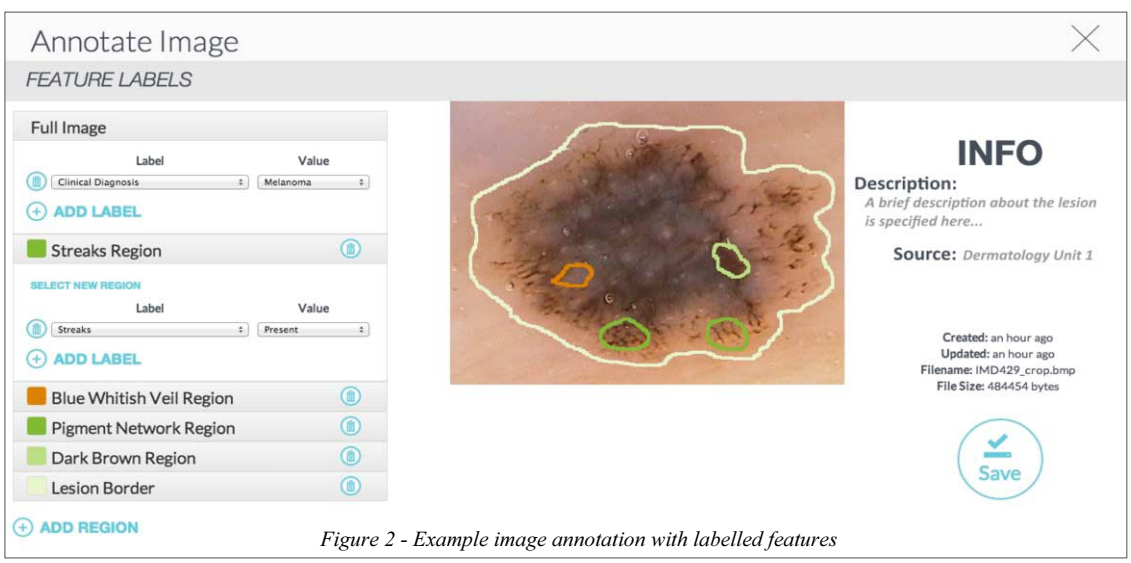


Figure 2 - Example image annotation with labelled features

can optionally fine-tune the analytic pipeline kernel and parameters provided to the training algorithms.

One feature of the system enables the domain expert to improve previously built models. After adding and annotating new samples to a previously trained dataset and model, the user can select "Update Model". This will trigger an incremental learning mechanism, which updates the existing classifier and builds an enhanced classification model without retraining the entire training set.

Classify

After building the model in the training phase, the clinician can now test the model on unseen images. Our system enables a fast classification (Figure 3) by uploading a dermoscopy image from the application. Using one or more classification models developed in the training stage, the system provides a diagnosis (e.g. melanoma vs. normal) and a confidence score on the classification result.

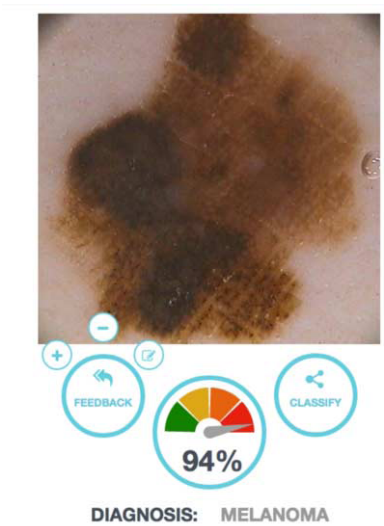


Figure 3 - Classify operation results and feedback controls

Feedback

To continually improve classifier performance, the system provides a mechanism to gather classification result feedback from expert users. As shown in Figure 3, after a classification is complete, the clinician can choose to provide feedback to the classifier score. Feedback can be provided in two ways. The simplest is with a binary answer, either agree or disagree with the diagnosis or score. The system also provides a third, structured feedback option. These feedback results are stored in the database and can be reviewed by an expert to help refine the classifier. The second option enables the user to "mark up" the classified image by providing a new annotation. In this process, the image annotation tool, shown in Figure 2, is used to define the annotation structure, including a clinical diagnosis. The previously untrained image, along with the new annotation, is then appended to the classifier model's existing training dataset and the classifier is incrementally retrained. We use stochastic gradient descent SVM [26, 27] as backend incremental learning method. When the feedback is available from expert, the already trained SVM is updated using gradient descent mechanism. Eventually, as more feedback is provided (more training examples), the SVM accuracy improves.

System and Infrastructure

The infrastructure component is implemented using cloud-computing principals of elastic scaling and deployment platform flexibility. The system architecture (Figure 4) is made up of four subcomponents, an application middleware providing consumable web services, a collection of worker processes for executing tasks on the analytics pipeline, a priority job broker for queuing and distributing tasks to workers, and an extensible document and media storage platform. Each component in the system is built to operate asynchronously and uses technologies, such as Web Sockets, to receive requests and provide an asynchronous response.

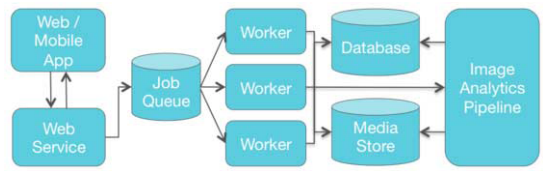


Figure 4 - Overview of system architecture

Web Service Interface

Access to the system is through web service application programming interfaces (APIs) using a representational state transfer (REST) methodology, JavaScript Object Notation (JSON) data format, and implemented in JavaScript using Node.js[2]. These APIs provide an authenticated web service to submit jobs for execution on the analytics pipeline. The system also provides a media management web service for uploading new media files for classification and training.

As many of the tasks are "long-running", in that they take more than three seconds to complete, the web services are designed to execute asynchronously. When a long-running request is submitted, the service will provide an immediate synchronous response with a task identifier then close the channel. The application has two options for receiving the acquiring response. First, the application can on-demand poll a result web service by providing the task identifier. Additionally, the application can subscribe to an asynchronous messaging channel using Web Sockets. In this situation, the web service will push a response to the application as soon as it becomes available.

Compute and Storage Infrastructure

Execution of tasks on the analytics pipeline is handled by a set of worker processes, implemented in Python, calling machine learning runtimes built on C, Java™, and Python code libraries. These workers provide an interface layer for accessing the analytics pipeline in a scalable and predictable manner. Individual workers subscribe and listen for waiting tasks from a job broker. When a worker process receives a job from the broker, it downloads required models and data from a media store, executes the analytic pipeline for a set of parameters (default or specified), and provides an asynchronous response back. The worker processes are pooled and stateless execution routines can be created and destroyed as needed.

The job broker monitors system load and provisions or de-provisions virtual machine (VM) instances as needed. For each instance, a set of worker processes are spawned to fill available capacity, taking into account available CPU, memory, and network capacity. We implement the job broker

2 http://nodejs.org/

process using Redis[3], to exclusively block and distribute waiting jobs, and Celery[4] for managing distributed workers on a set of VMs.

The system supports several storage options. For metadata storage and management, the system uses a document store, such as MongoDB[5]. For large, opaque binary data, including media (images) and classifier models, the system supports multiple data stores. In addition to a MongoDB-based storage backend, the system also supports distributed object stores, such as Amazon Simple Storage Service (S3)[6] and OpenStack Swift[7].

While the system can operate in public cloud service platforms, such as IBM BlueMix[8] or Amazon Elastic Compute Cloud (EC2)[5], it is designed to support a hybrid of public and private infrastructure. As many countries require personal health data to remain onshore [24], this separates web service APIs and applications from data storage and computation, simplifying security and monitoring requirements. In addition to the regulatory considerations, this offers other advantages. In some situations, it may be inefficient or cost prohibitive to transfer and store certain types of media in cloud object or block stores [25]. By keeping data on-site or by providing a "bring your own" storage environment, data storage as well as billing and payment is outsourced to other services or local infrastructure, further reducing security and latency concerns.

Results and Discussion

Annotation

Traditionally, annotating images for input into training modules is done using a spreadsheet tool such as Excel [28]. In tests, the time to annotate a series of images using our system's annotate image tool is significantly faster, especially for images that require drawn annotations. Despite these enhancements, we found the process to still be extremely time consuming. Future versions of the annotation tool will improve the annotation process with contextually aware user interface enhancements along with a feedback-trained "smart" annotation tool. The smart annotator will suggest an annotation structure for a previously untrained image and allow the clinician to accept or amend the suggestion. These modifications provide further feedback to future annotate and train operations.

Training and Feedback

The feedback mechanism of the system enables incremental updates to the classification model as updates arrive over time. Figure 5 demonstrates an example of the accuracy of the classification model as new user-provided feedback responses (either accepting the prediction or rejecting it) are received over time. The example set contains both "agreement" and "disagreement" from the experts. The graph suggests that the classifier struggles to predict correctly for the first few samples, but, as the system receives more feedback responses, the overall accuracy of the model improves. The initial classification model is trained using one example, then gradually updated by receiving more training samples. In this example, the classification accuracy for the first ten feedback responses is 56% but quickly improves to 75% over the next fifteen. The accuracy continues to improve with some fluctuation before converging at approximately 90% after 150 responses.

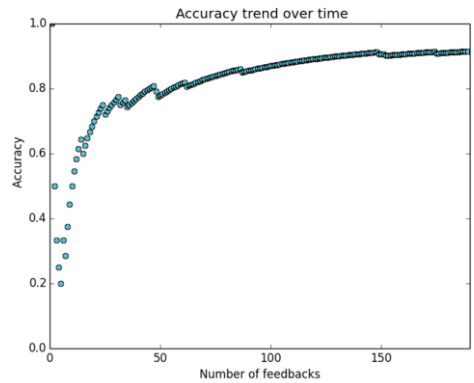


Figure 5 - Accuracy of classify operations as number of feedback results increases

Classification and Scalability

To test classifier performance and response time, we built a new classifier model from a corpus of approximately 100 images. We selected five images of similar size and resolution for classifier testing. Three images were of positive diagnosis (melanoma), two negative (normal skin). The five images were randomly selected and submitted in variable time increments to the web service, which submitted the images as jobs to the job broker. The job broker and worker processes were provisioned onto three virtual machines (VMs), each providing eight compute cores and 16 gigabytes of memory.

For each classification operation, the baseline response time was approximately 6.5 seconds per task (no additional system load), with an image upload time of 0.5 - 0.75s and classifier runtime of 5.7 - 6.0s. As expected, by varying the number of concurrent worker processes per VM, we found the classifier jobs scale with the number of compute cores linearly, indicating efficient CPU utilization by the classifier. Increasing the number of concurrent jobs beyond the number of compute cores for the VM resulted in an increase in classification and queue time. Additional work is needed to model and optimise total wait time and classification time against the costs associated with available but unused compute resources.

Conclusions

In this paper, we have described an innovative feedback-driven system for performing melanoma recognition in dermoscopy images. This system provides domain experts (clinicians) easy access to the complexities of a machine learning system, including annotation, training, and classification, through a simple user interface. Additionally, the system is built on top of a cloud-enabled infrastructure that can scale to thousands of users. Next steps involve further work to refine the training annotation workflow and a public deployment with one or more clinical partners.

3 http://redis.io/
4 http://www.celeryproject.org/
5 http://www.mongodb.org/
6 http://aws.amazon.com/
7 http://swift.openstack.org/
8 https://console.ng.bluemix.net/

Acknowledgments

Thanks to the PH^2 project for providing the images used in our model development, testing, and figures [28].

References

[1] Hey T, Trefethen A. The Data Deluge: An e-science perspective. In: Berman F, Fox G and Hey T, eds. Grid Computing: Making the global Infrastructure a Reality. John Wiley and Sons, 2003.

[2] Kononenko I. Machine learning for medical diagnosis: history, state-of-the-art and perspective. Artificial Intelligence in Medicine. 2001;23(1):89-109.

[3] Gammerman A. Modern machine learning techniques and their applications to medical diagnostics. In: Papadopoulos H, Andreou AS, and Bramer M. Artificial intelligence applications and innovations. Springer Link, 2010.

[4] Wang S, Summers RM. Machine learning and radiology. Medical Image Analysis. 2012;16(5):933-951.

[5] Fujita H, Nogata F, Jiang H, Kido, S. Medical image processing and computer-aided detection/diagnosis (CAD). Proc Intl Conf Computerized Healthcare (ICCH). 2012;66-71.

[6] Doi K. Computer aided diagnosis in medical imaging: Historical review, current status and future potential. Computer-aided Diagnosis and Image-guided decision support. 2007;31(4-5):198-211.

[7] Sajda P. Machine learning for detection and diagnosis of disease. Ann Rev of Biomedical Engineering. 2007;8:537-565.

[8] Rolim CO, Koch FL, Westphall CB, Werner J, Fracalossi A, Salvador GS. A cloud computing solution for patient's data collection in health care institutions. Proc Intl Conf E-Health, Telemedicine and Social Medicine. 2010;95-99.

[9] Kuo AM. Opportunities and challenges of cloud computing to improve health care services. J Med Internet Res. 2011;13(3).

[10]Zhang R, Liu L. Security models and requirements for healthcare applications clouds. Proc Intl Conf Cloud Computing. 2010;268-275.

[11]Avila-Garcia MS, Trefethen AE, Brady M, Gleeson F, Goodman D. Lowering the barriers to cancer imaging. Proc Intl Conf Science. 2008;63-70.

[12]Shaozhen Y, Hanbing W, Yubin C. Design for medical imaging services platform based on cloud computing technologies. Proc Intl Conf Cloud Computing and Big Data. 2013;455-460.

[13]McGregor C. A cloud computing framework for real-time rural and remote service of critical care. Proc Intl Symp Computer based medical systems. 2011;1-6.

[14]Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. Comm of the ACM. 2008;51(1):107-113.

[15]Low Y, Gonzales J, Kyrola A, Bickson D, Guestrin C, Hallerstein JM. Distributed GraphLab: A framework for machine learning and data mining in cloud. Proc VLDB Endowment. 2012;5(8):716-727.

[16]Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical text analysis and knowledge extraction systems (cTAKES): architecture, component, evaluation and applications. J Amer Med Inform Assoc. 2010;17:507-513.

[17]O'Drsicoll A, Daugelaite J, Sleator RD. 'Big Data', Hadoop and cloud computing in genomics. J Biomedical Informatics. 2013;46(5):774-781.

[18]Kotsiantis SB. Supervised Machine Learning: A review of Classification Techniques, In Emerging Artificial Intelligence Applications in Computer Engineering. IOS Press, 2007.

[19]Ziou D, Tabbone S. Edge detection techniques: An overview. Intl J Pattern Recog and Image Analysis. 1998; 8(4):537–559.

[20]Russ JC. The Image Processing Handbook. Sixth ed. CRC Press, 2011.

[21]Pehamberger H, Steiner A, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. I: Pattern analysis of pigmented skin lesions. J American Academy of Dermatology. 1987;17:571–583.

[22]Stolz W, Riemann A, Cognetta AB, Pillet L, Abmayr W, Holzel D, Bilek P, Nachbar F, Landthaler M, Braun-Falco O. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. Euro J Dermatology. 1994;4:521–527.

[23]Henning J, Dusza S, Wang S, Marghoob A, Rabinovitz H, Polsky D, Kopf A. The CASH (colour, architecture, symmetry, and homogeneity) algorithm for dermoscopy. J Amer Acad Dermatology. 2007;56(1):45–52.

[24]Kuo AM-H. Opportunities and Challenges of Cloud Computing to Improve Health Care Services. J Med Internet Res. 2011:13(3):e67. doi:10.2196/jmir.1867.

[25]Symantec. Avoiding the Hidden Costs of the Cloud, 2013. Available from: http://www.symantec.com /content/en/us/about/media/pdfs/b-state-of-cloud-global-results-2013.en-us.pdf

[26]Ruppert D. Efficient estimations from a slowly convergent Robbins-Monro process. Technical Report 781, Cornell University Operations Research and Industrial Engineering.1988.

[27]Shalev-Shwartz S, Singer Y, Srebro N. Pegasos. Primal Estimated subGrAdient sOlver for Svm. Proc of the 24th Intl Conf Machine Learning. 2007.

[28]Mendonça T, Ferreira P, Marques J, Marcal A, Rozeira J. PH^2 - A dermoscopic image database for research and benchmarking. Proc Intl Conf IEEE Engr Medicine and Biology Society. 2013.

Address for correspondence

Rahil Garnavi, IBM Research – Australia, Level 5, 204 Lygon Street, Carlton, VIC, 3053, Australia, rahilgar@au1.ibm.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-696

A Two-stage Dynamic Model to Enable Updating of Clinical Risk Prediction from Longitudinal Health Record Data: Illustrated with Kidney Function

Artur Akbarov[a], Richard Williams[a,c], Benjamin Brown[a,c], Mamas Mamas[a,b], Niels Peek[a,c], Iain Buchan[a,c], Matthew Sperrin[a]

[a] *Health eResearch Centre, Farr Institute, University of Manchester, UK*
[b] *Institute of Cardiov ascular Sciences, University of Manchester, UK*
[c] *Greater Manchester Primary Care Patient Safety Translational Research Centre, Universtity of Manchester, UK*

Abstract

We demonstrate the use of electronic records and repeated measures of risk factors therein, to enable deeper understanding of the relationship between the full longitudinal trajectory of risk factors and outcomes. To illustrate, dynamic mixed effect modelling is used to summarise the level, trend and monitoring intensity of kidney function. The output from this model then forms covariates for a recurrent event Cox proportional hazards model for predicting adverse events (AE). Using data from Salford, UK, our multivariate model finds that steeper declines in kidney function raise the hazard of AE (HR: 1.13, 95% CI (1.05, 1.22)). There is a non-proportional relationship between the hazard of AE and the monitoring intensity of kidney function. Neither of these variables would be present in a classical risk prediction model.. This work illustrates the potential of using the full longitudinal profile of risk factors, rather than just their level. There is an opportunity for deep statistical learning leading to rich clinical insight using longitudinal signals in electronic data.

Keywords:

Biomarkers; Clinical Prediction; Estimated Glomerular Filtration Rate; Longitudinal Analysis; Model Calibration.

Introduction

Most risk prediction models incorporate only the most recently measured value of a biomarker or risk factor. This is often inadequate as risk may be driven by decline, variability, or other complex aspects of the longitudinal trajectory. Electronic medical records contain rich longitudinal data from which risk factor trajectories can be estimated, and therefore used to enrich prediction models. In this paper we present an example of a two-stage model that first estimates the longitudinal trajectory of a risk factor, then uses predictors derived from this in a risk prediction model.

We illustrate the model using the example risk factor of estimated glomerular filtration rate (eGFR), which is a measure of kidney function. Deteriorating kidney function is an important risk factor for adverse events. Chronic heart failure (CHF) and type 2 diabetes (T2DM) both damage the kidneys. CHF reduces blood flow through the kidneys due to weaker heart pumping; the factors causing it also affect the kidney blood vessels; and some CHF treatments may damage the kidney. T2DM damages the blood vessels and other parts of the kidney directly. So, patients with these conditions require particularly diligent monitoring of their regular laboratory test results.

Kidney disease in patients with T2DM is well documented in the literature. A recent study in the US showed that as many as 43.5% of patients with T2DM have chronic kidney disease (CKD) [1]; CKD is strongly associated with increased mortality and accelerated cardiovascular disease [2]. Patients with T2DM are also more likely to experience acute renal failure than patients without diabetes (adjusted hazard ratio: 2.5, 95% CI 2.2 – 2.7) [3]. Like CKD, acute renal failure is also associated with high mortality rate, around 50% [4].

Data from the EuroHeart Failure survey suggest that patients with T2DM and impaired kidney function have amongst the worst short and long-term outcomes [5]. A recent study [6], showed that the combination of poor heart and kidney function is a particularly strong and discrete risk factor for death. Another study [7] found that T2DM was a statistically significant predictor of all-cause mortality in patients with CHF, but only those who had eGFR between 30 and 90 ml/min*1.73m^2. Heart and kidney dysfunction are closely linked [8] and diabetes worsens the situation, whether measured by hospital admissions or death rates [9].

In addition to slowly progressing disease, the kidneys can suffer acute injury leading to emergency hospitalisation and death, especially among patients with CHF and T2DM. Most reported studies in this area have focused on measuring kidney function at the time of endpoints such as death or admission. This study shifts the focus to the *process* of deteriorating kidney function itself. The aim is to examine in detail the trajectory of eGFR prior to endpoints being observed. In particular we will consider the effects of eGFR level, rate of decline and intensity of monitoring on adverse events such as emergency hospitalisation rates and all-cause mortality in patients with CHF and T2DM.

Methods

The data used in this study were obtained from the Salford Integrated Record (SIR) – a medical data store in Salford, Greater Manchester, UK. The population of Salford is around 234 thousand with 53 general practices and a single hospital. SIR contains linked data from both primary and secondary care for the population that engage with health services.

The cohort was restricted to patients who had both CHF and T2DM. Patients with congenital heart disease and those with cor pulmonale were excluded from the study, as were patients who had renal replacement therapy during the study period, where eGFR has a different interpretation. The study period spanned from January 2008 to August 2012.

The main outcome of interest was adverse events (AE), which comprise emergency hospital admissions and all-cause mortality. The covariates of primary interest were eGFR level, rate of decline and intensity of monitoring. Models were corrected for patient's age at the time of follow-up start, gender, and index of multiple deprivation (IMD) 2010. All covariates related to

eGFR are time dependent and as such require estimation of the longitudinal process for each patient. Therefore we take a two stage modelling approach. In Stage 1, the eGFR process is estimated. Stage 2 then models AEs as a function of the eGFR process.

For Stage 1, a simple approach to estimating the eGFR level at any time, t, is to use either the most recent eGFR value or some summary of a collection of recent eGFR values like the median or the mean. The eGFR decline can be calculated as the difference between the last and the penultimate eGFR values. In both cases, one needs to make several, rather arbitrary, decisions, for example, decide how far apart the two eGFR values should be in the case of calculating the eGFR decline. Also, as eGFR values can be 'noisy', this approach can be affected by variations in eGFR values. In order to smooth the variation in eGFR values and use all of the information about the degradation process of the kidney function in this particular population, we used dynamic mixed effects regression [10-11]. The eGFR level and decline for each patient at any given point in time t, was derived from a mixed effects model fitted to all of the eGFR data prior to t. The generic model was formulated as follows: dependent variable – the logarithm of eGFR, predictors (fixed effects) – age, gender, IMD, time since diagnosis (TSD) and time since study start (TSS), and random effects – variance of random effects on the intercept and coefficient of TSS.

For each patient at each endpoint at time t, eGFR level was measured by the predicted eGFR value at time t and eGFR decline was measured as $(a - b) / (t - t_a)$, where a is the predicted eGFR value at the start of the study, t_a, and b is the predicted eGFR value at time t. Since the kidney function for the cohort considered in this study is in progressive decline, for all $t > t_a$ the eGFR slope is a positive number, where higher values indicate steeper decline.

The eGFR intensity of monitoring was measured as the proximity of the most recent eGFR record: within 180 days, between 180 and 365 days and more than 365 days.

For Stage 2, modelling the risk of AE, we used a repeated measures Cox proportional hazards (CPH) model with gamma shared frailty – i.e. patient level random effects were modelled using a gamma distribution [12]. CPH model is a popular method used to model survival data [13-14]. It is a semi-parametric method with an advantage over parametric methods in that it focuses on hazard ratios only, thus, bypassing the need to specify a parametric distribution for the baseline hazard. This allows direct comparison of the risk between subjects with different characteristics, where a hazard ratio of one indicates that there is no difference in risk; ratios above one indicate higher risk; and below one lower risk.

Results

The process of deriving the cohort is illustrated in Figure 1.

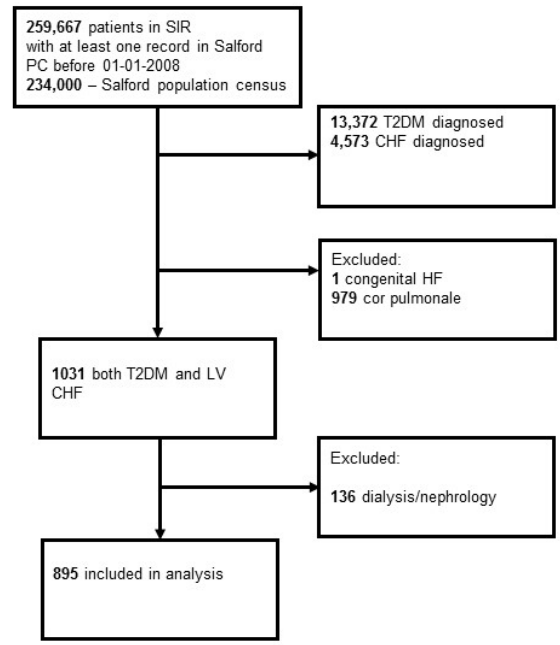


Figure 1 – Exclusion flowchart.

Stage 1

Table 1 shows exponentiated fixed effects of two log eGFR mixed effects models. The first model is the model fitted for the first AE during the study period. It is the model that used the smallest number of eGFR records for estimating the kidney function decline process. For each subsequent event, this model was updated with more data until the last model, which was fitted at the end of the study. Time is measured in years. From Table 1 we can note that older patients tend to have lower eGFR, female patients have lower eGFR than male patients, patients who have had both CHF and T2DM for longer have lower eGFR, and the eGFR is in decline with time.

Figure 2 shows the random effects of each patient on the intercept and the coefficient of the TSS for the last model. For the intercept, patients with negative random effects have lower than average eGFR. For the slope, patients with negative random effects have steeper than average decline in eGFR. We can see from Figure 2 that most of the steeper decline in eGFR is associated with patients with below average eGFR. The predicted eGFR levels and declines where carried forward to Stage 2.

Table 1 – Exponentiated parameter estimates of log eGFR models. Interpretation example: male coefficient is greater than one, meaning male eGFR is predicted to be higher.

	First model	**Last model**
Intercept	139.5 (119.72, 162.56)[a]	146.01 (126.47, 168.56)[a]
Age	0.99 (0.99, 0.99)[a]	0.99 (0.99, 0.99)[a]
Gender:		
Female	1	1
Male	1.07 (1.03, 1.11)[a]	1.05 (1.01, 1.1)[a]
Time since diagnosis	0.98 (0.98, 0.99)[a]	0.98 (0.98, 0.99)[a]
Time since study start	0.96 (0.95, 0.98)[a]	0.97 (0.97, 0.98)[a]

Key: a – statistically significant at 5%

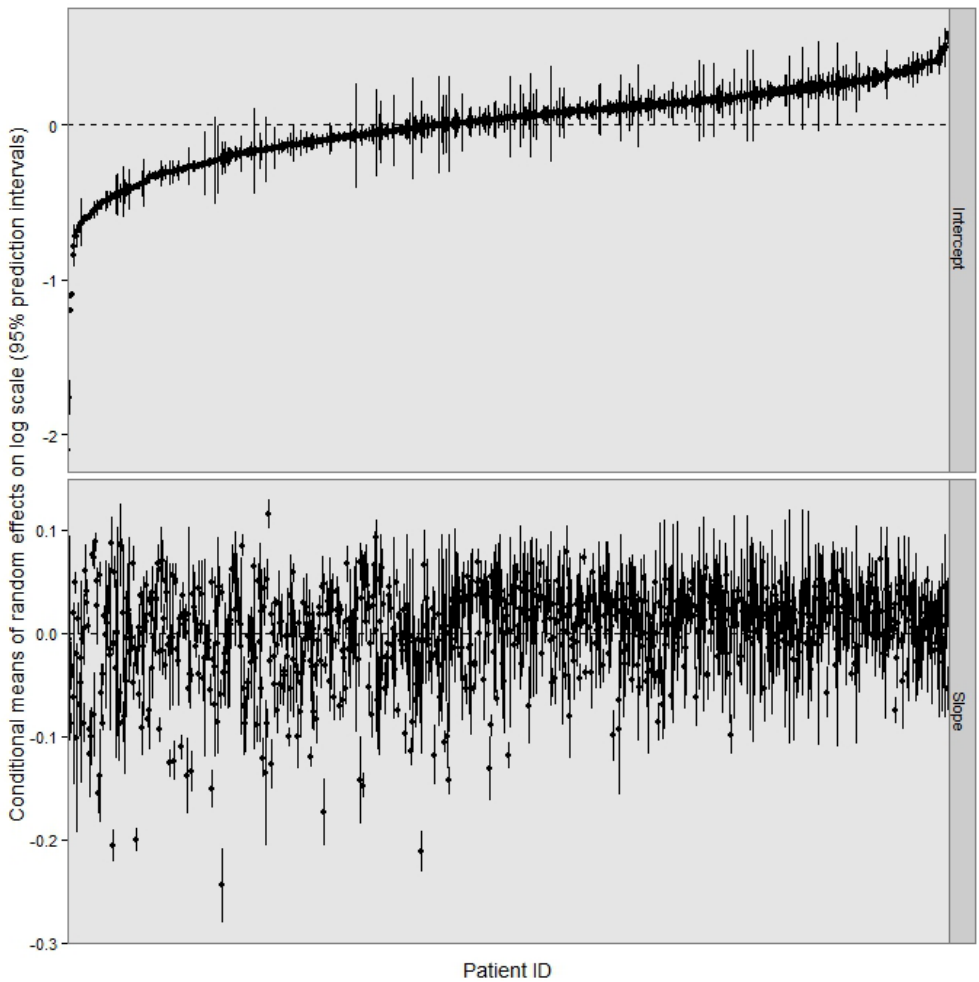


Figure 2 – Patient random effects on the intercept and on the slope of time since study start [15].

Stage 2

Table 2 shows covariate details and hazard ratios for univariate and multivariate models for AE. We can note the following from the table. Older patients are at higher risk of AE. Gender, IMD and eGFR level were not statistically significant, whereas the eGFR decline was. The hazard ratio for the eGFR decline indicates that a one unit decrease in eGFR per year increases the hazard of AE by 10-11%. The hazard ratios for the eGFR monitoring indicate that patients with more recent eGFR records are more likely to have AE. The multivariate model was selected on the basis of significance of the parameter estimates. The eGFR monitoring covariate failed to adhere to the proportional hazards assumption, so the multivariate model is stratified on this. Kaplan-Meier plots of the eGFR monitoring strata are presented in Figure 3.

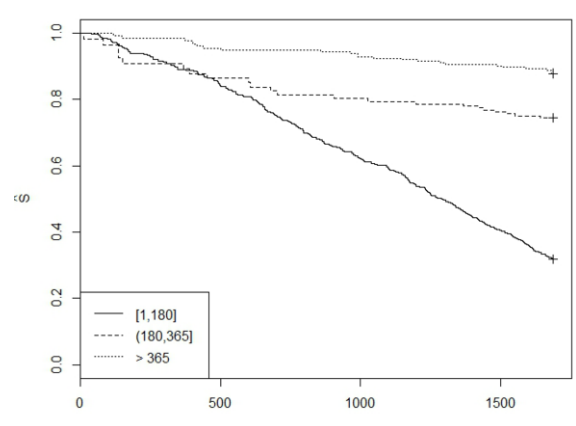


Figure 3 – Kaplan Meier plot of AE-free survival stratified by time since most recent eGFR measurement.

Table 2 – Hazard ratios for emergency hospital admissions models. Hazard ratios greater than one represent higher risk.

Covariates	**Mean (SD), count or proportion**	**Hazard ratio (95% CI)**	
		Univariate	**Multivariate**
Age	73.03 (10.43)	1.02 (1.01, 1.04)[a]	1.03 (1.01, 1.05)[a]
Gender:			
Female	42%	1 (reference)	-
Male	58%	1.2 (0.857, 1.68)	-
IMD 2010	37.16 (18.20)	1 (0.993, 1.01)	-
eGFR level	59.03 (18.18)	0.996 (0.988, 1)	-
eGFR decline	2.27 (2.05)	1.11 (1.03, 1.19)[a]	1.13 (1.05, 1.22)[a]
Last eGFR	58.58 (21.37)	0.993 (0.986, 1)[a]	-
eGFR monitoring:			
[*1, 180*]	730	1	-
(*180, 365*]	149	0.21 (0.12, 0.35)[a]	-
> 365	190	0.13 (0.08, 0.20)[a]	-

Key: a – statistically significant at 5%

Discussion

We have presented a proof-of-concept study of a two-stage framework for modelling the full longitudinal profile of a risk factor and its impact on risk, using the example of eGFR decline in T2DM and CHF patients. We found that eGFR decline, modelled as patient-specific slope from a mixed effects model, was a strong predictor for adverse events (AE), and eGFR level was not a significant predictor either in a univariate or multivariate model. In a standard predictive modelling framework, only the most recent eGFR measurement would have been taken, so eGFR decline would not have been captured. We additionally considered the intensity of eGFR measurement as a predictor, and found a complex relationship between this and AE.

The main strength of this study is the incorporation of longitudinal information from routine care-records into risk prediction, a feature absent from standard models. We considered one risk factor but the method is easily extended to multiple risk factors. The main weakness is that uncertainty in eGFR predictions is not propagated to the estimates of the hazard ratios across the two stages of the method. Future work will investigate joint modelling approaches to overcome this.

We considered only mixed effects regression to model the trajectory of eGFR. Other methods such as random forests, neural networks and support vector regression could be used for this purpose. These machine-learning techniques might better cope with additional covariates, particularly treatments for T2DM and CHF that could influence kidney function. In our data, there were over a thousand different medications prescribed among our patient cohort in the month preceding the first eGFR measurement. Future work will compare alternative dynamic modelling strategies capable of incorporating complex covariate structures.

The ability to monitor and predict longitudinal processes using dynamic models, integrated with medical records, provides an opportunity to develop optimal monitoring polices for these risk factors. Linking the predictions of the kidney function decline to the effectiveness of its monitoring, aimed at averting adverse events, is a hard problem and the subject of ongoing research. This requires investigating the mechanism behind how the decline in eGFR leads to adverse events. For example, it may be that the decline in eGFR first leads to changes in treatment that in turn precipitate adverse events.

As with most routine care-records, the data in our study had limitations. In particular, the linkage of emergency hospital admissions to the SIR database required general practitioners to enter or accept data on admission episodes. In addition, we did not have data on the duration of hospital admissions, which could have been used to refine our endpoint definitions.

Despite limitations in the data and methodology this study has clearly demonstrated that clinical predictive models can improve their predictive abilities by taking account of the longitudinal information in electronic health records.

Consider EU Directive 2007/47 that considers algorithms used for prognostic purposes to be “medical devices.” The usual validation methods for such algorithms certify the encoding into software of models, but rarely question the fact that published model parameters will drift in calibration as risk-factor environments change. In addition, the structure of models may need reconsidering as populations, contexts of healthcare/observation and data-quality evolve.

Conclusion

Sophisticated statistical modelling with electronic health records allows for complex relationships between risk factor trajectories and outcomes to be captured, thereby improving predictions. Most current predictive models do not exploit longitudinal health record information in this way – neither is the regulatory framework for clinical algorithms set up to cope with this dynamic modelling approach. Research is urgently needed into statistical methods and computational frameworks for optimising clinical prediction from the continuously evolving risk information across heterogeneous populations, healthcare environments and information systems.

Acknowledgements

This work was supported by the University of Manchester's Health eResearch Centre (HeRC) funded by the Medical Re-

search Council Grant MR/K006665/1; and by the National Institute for Health Research Greater Manchester Primary Care Patient Safety Translational Research Centre (NIHR GM PSTRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

[1] Bailey RA, Wang Y, Zhu V, and Rupnow MFT. Chronic kidney disease in US adults with type 2 diabetes: an updated national estimate of prevalence based on Kidney Disease: Improving Global Outcomes (KDIGO) staging, *BMC Research Notes*, 2014: 7 (415).

[2] Sarnak MJ, Levey AS, Schoolwerth AC, Coresh J, Culleton B, Hamm LL, McCullough PA, Kasiske BL, Kelepouris E, Klag MJ, Parfrey P, Pfeffer M, Raij L, Spinosa DJ and Wilson PW; American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention, Hypertension, 2003: 42 (5), pp. 1050-1066.

[3] Girman CJ, Kou TD, Brodovicz K, Alexander CM, O'Neill EA, Engel S, Williams-Herman DE, and Katz L. Risk of acute renal failure in patients with Type 2 diabetes mellitus, Diabetic Medicine. 2012: 29: pp. 614-621.

[4] Ympa YP, Sakr Y, Reinhart K, and Vincent JL. Has mortality from acute renal failure decreased? A systematic review of the literature, The American Journal of Medicine. 2005: 118: pp. 827-832.

[5] Harjola VP, Follath F, Nieminen MS, Brutsaert D, Dickstein K, Drexler H, Hochadel M, Komajda M, Lopez-Sendon JL, Ponikowski P and Tavazzi, L. Characteristics, outcomes, and predictors of mortality at 3 months and 1 year in patients hospitalized for acute heart failure. European Journal of Heart Failure, 2010: 12(3): pp. 239-248.

[6] Cioffi G, Rossi A, Targher G, Zoppini G, de Simone G, Devereux RB, Bonora E, and Vassanelli C. Usefulness of subclinical left ventricular midwall dysfunction to predict cardiovascular mortality in patients with type 2 diabetes mellitus, American Journal of Cardiology. 2014: 113: pp. 1409-1414.

[7] Russo G, Cioffi G, Tarantini L, Cherubini A, Faganello G, Mazzone C, Barbati G, Candido R, Faggiano P and Di Lenarda A. Does renal function influence the prognostic impact of type 2 diabetes mellitus in patients with chronic heart failure and left ventricular dysfunction?, IJC Metabolic & Endocrine. 2014: 4: pp. 53-57.

[8] Liu Y, Gao L, Xue Q, Yan M, Chen P, Wang Y, and Li Y.. Impact of renal dysfunction on long-term outcomes of elderly patients with acute coronary syndrome: a longitudinal, prospective observational study, BMC Nephrology. 2014: 15 (78).

[9] Carrasco-Sánchez FJ, Gomez-Huelgas R, Formiga F, Conde-Martel A, Trullàs JC, Bettencourt P, Arévalo-Lorido JC, and Pérez-Barquero MM. Association between type-2 diabetes mellitus and post-discharge outcomes in heart failure patients: Findings from the RICA registry, Diabetes Research and Clinical Practice. 2014: 104: pp. 410-419.

[10] Wickham H. The split-apply-combine strategy for data analysis. Journal of Statistical Software. 2011: 40 (1): pp. 1-29. http://www.jstatsoft.org/v40/i01/.

[11] Bates D, Maechler M, Bolker B, and Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. http://CRAN.R-project.org/package=lme4. 2014.

[12] Therneau T. A package for survival analysis in S. R package version 2.37-4, http://CRAN.R-project.org/package=survival. 2014.

[13] Cox, D. R. Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society, Series B 34: pp. 187—220. 1972.

[14] Kalbfleish, J. D. and R. L. Prentice (1980). The statistical analysis of failure time data. New York: John Wiley. 1980.

[15] Wickham H. ggplot2: elegant graphics for data analysis. Springer New York. 2009.

Address for correspondence

Matthew Sperrin, Health eResearch Centre, Farr Institute, Vaughan House, Portsmouth St, University of Manchester, M13 9PL. Email: matthew.sperrin@manchester.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-701

Development and preliminary validation of a dynamic, patient-tailored method to detect abnormal laboratory test results

Paolo Fraccaro[a,b,c], Benjamin Brown[a,b,c], Mattia Prosperi[b,c], Matthew Sperrin[b,c], Iain Buchan[a,b,c], Niels Peek[a,b,c]

[a] *NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, University of Manchester, Manchester, UK.*
[b] *Health eResearch Centre, Farr Institute for Health Informatics Research, UK.*
[c] *Centre for Health Informatics, University of Manchester, Manchester, UK.*

Abstract

Laboratory test results in primary care are flagged as 'abnormal' when they fall outside a population-based Reference Interval (RI), typically generating many alerts with a low specificity. In order to decrease alert frequency while retaining clinical relevance, we developed a method to assess dynamic, patient-tailored RIs based on mixed-effects linear regression models. Potassium test results from primary care were used as proof-of-concept test bed. Clinical relevance was assessed via a survey administered to general practitioners (GPs). Overall, the dynamic, patient-tailored method and the combination of both methods flagged 20% and 36% fewer values as abnormal than the population-based method. Nineteen out of 43 invited GPs (44%) completed the survey. The population-based method yielded a better sensitivity than the patient-tailored and the combined methods (0.51 vs 0.41 and 0.38, respectively) but a lower PPV (0.66 vs 0.67 and 0.76, respectively). We conclude that a combination of population-based and patient-tailored RIs can improve the detection of abnormal laboratory results. We suggest that lab values outside both RIs be flagged with high priority in clinical practice.

Keywords:

Clinical Decision Support Systems; Point-of-Care Systems; Biochemistry.

Introduction

Failure to follow up laboratory test results is a major concern in primary care [1]. Missed test results can lead to delayed interventions or to decisions made on the basis of incomplete information with potential compromise of patient safety [2]. For patients, poor test handling may generate increased number of visits, repeated laboratory examinations, and unnecessary stress or harm [3]. Callen *et al.* [1] showed that the extent of the failure to follow up test results in primary care ranges from 7% to 62% of results and can cause delayed diagnosis, preventable hospitalisation and adverse drug reactions.

Electronic Health Records (EHRs) can improve test result follow-up and management [4]. However, general practicioners' (GPs) satisfaction of informatics systems in place to manage tests is low [5]. One of the main barriers for timely follow-up is the high number of often unecessary alerts that physicians are presented with on a daily basis [6, 7]. It is estimated that each general practitioner spends almost an hour per day processing alerts generated by primary care information systems, the majority of which are composed of test result alerts [5, 6]. As a consequence, physicians may not have the necessary time to focus on the most important alerts [8]. This information overload contributes to alert fatigue [9] and can potentially generate patient safety issues [1].

The high number of alerts produced is directly related to the lack of specificity of current threshold-based methods for detecting abnormal values. The majority of EHRs use population-based Reference Intervals (RIs), which are defined as "intervals that, when applied to the population serviced by the laboratory correctly include most of the subjects with characteristics similar to the reference group and exclude the others" [10]. RIs are usually calculated by assuming a Gaussian distribution of test results for the given physiological measure (e.g. potassium, creatinine, heamoglobin), and estimating population mean and variance to calculate the RI as the 95% reference range. By definition a value outside the RI is flagged as abnormal [11]. As this approach is based on population estimates, it may flag values as abnormal that may be considered normal in the context of a specific patient's medical history. For instance, some patients have persistently high or low levels of certain physiological parameters, and this is dealt with via the clinician's knowledge of the patient. Conversely, some patients may experience sudden changes in critical parameters that need prompt medical action, but these changes are not flagged as abnormal because measured values are still within the population-based RI. In this regard, the adoption of more personalised methods that adapt to the patient's history could reveal key patterns and insights the interpretation based on population thresholds would not [12].

The aim of the paper is to develop a dynamic patient-tailored method for detecting abnormal laboratory test results in primary care, and assess the incidence of abnormal test results when using this method alone and in combination with the population-based method. Furthermore, we aimed to assess the potential clinical relevance through a survey administred to experienced practicioners in UK primary care.

Methods

Model

Mixed-effects modelling allows the analysis of data with complex patterns of variability and hierarchical structure [13]. Specifically, when estimating RI for laboratory tests, mixed-effects modelling can take into account both the population variability and the intra-patient variability. Consider a patient i and its corresponding test result set $Y_i=y_{i1}..$, by applying the introduced hierarchical structure and calculating the maximum likelihood estimate:

If $y_{ij} \sim N(\alpha_i, \sigma^2)$, $\alpha_i \sim N(\mu,\omega^2)$, then $\alpha_i \mid \bar{y}_{ij}, \sigma^2, \mu,\omega^2 \sim N(\tilde{\mu}_{ij}, V_{ij})$ where

$$\tilde{\mu}_{ij} = \frac{\mu\omega^{-2} + \bar{y}_{ij}\frac{n_{ij}}{\sigma^2}}{\omega^{-2} + \frac{n_{ij}}{\sigma^2}} \quad \text{and} \quad V_{ij} = (\omega^{-2} + \frac{n_{ij}}{\sigma^2})^{-1}$$

Equivalently,

$$\tilde{\mu}_{ij} = \mu\lambda_{ij} + (1 - \lambda_{ij})\bar{y}_{ij} \qquad (1)$$

$$\lambda_{ij} = \frac{\omega^{-2}}{\omega^{-2} + \frac{n_{ij}}{\sigma^2}} = \frac{\frac{\sigma^2}{n_{ij}}}{\omega^2 + \frac{\sigma^2}{n_{ij}}} = \frac{V_{ij}}{V_{ij} + \omega^2} \qquad (2)$$

Here, μ and ω^2 are population mean and variance, y_{ij} is the jth observation of patient i, α_i is the mean of patient i, σ^2 is the intra-patient variance. Moreover, $\bar{y}_{ij}$ and n_{ij} are the sample mean and number of observations for patient i after j observations. Finally, $\tilde{\mu}_{ij}$ and V_{ij} are the maximum likelihood estimates of α_i and σ^2, and λ_{ij} is a shrinkage factor, that as soon as there are more observations for patient i, increases the weight of the sample mean $\bar{y}_{ij}$ compared to the population mean μ. Accordingly, by using $\tilde{\mu}_{ij}$ and V_{ij} it is possible calculate an adaptive patient-tailored RI for patient i (defined as in the standard method as the 95% reference range), which is dynamically updated at each new test result and tells us what is normal/abnormal in a specific patient's context. Figure 1 shows an example of the application of the population-based and patient-tailored RIs to a patient's potassium results time series to detect abnormal values. As it is possible to notice at the first observation (on the left), with no previous information about the spefic patient, the population-based and patient tailored RIs coincide. Conversely, with new observations coming the patient-tailored RI is updated and adapts to the the patient's context. For example, the third value, that is similar to the previous ones but slighlty outside the population-based RI, is not flagged as abnormal according the patient-tailored method but it is according to the population-based ones. Finally, the last value, that is inside the RI but "unusual" for the patient, is flagged as abnormal by following the patient-tailored RI and it would be missed with the population-based one.

Data

The Salford Integrated Record (SIR) database was used to test our method. SIR is an anonymized electronic health record database from the City of Salford (population 234k, UK), which collects data from 49 primary care providers and one secondary care provider. Records are stored using Read codes v2 and CTV3 standards [14].

We focused our analysis on potassium results and we extracted data from SIR for patients aged 18 to 85 during the period 1/1/1990–31/12/2012. Potassium results are influenced by medications and chronic conditions, and in case of hypokalemia (<2.5 mmol/l) or hyperkalemia (>6 mmol/l) it can result in significant cardiac dysrhythmias. In addition, some patients can have levels persistently outside the RI without any effect on heart ryhthm. Accordingly, a patient-tailored alerting system would be particularly helpful for this laboratory parameter. As a consensus RI across UK laboratories, we implemented in our study a potassium RI of 3.5 to 5 mmol/l [15].

Parameter estimation

From SIR data extraction, after outliers exclusion in order to reproduce the implemented potassium RI, we derived two datasets: 1) a test set made by 500 patients randomly exracted, with all their potassium results; 2) a training set composed by all the remaining data extract used for fitting the mixed-effects model and estimate its parameters to implement the abovementioned adaptive patient-tailored RI.

Potassium results in the test set were flagged as abnormal if out of the RI range according either to the standard threshold-based method using our UK consensus RI (referred from now on as *standard method*) or the adaptive patient-tailored RI *(patient-tailored method)*.

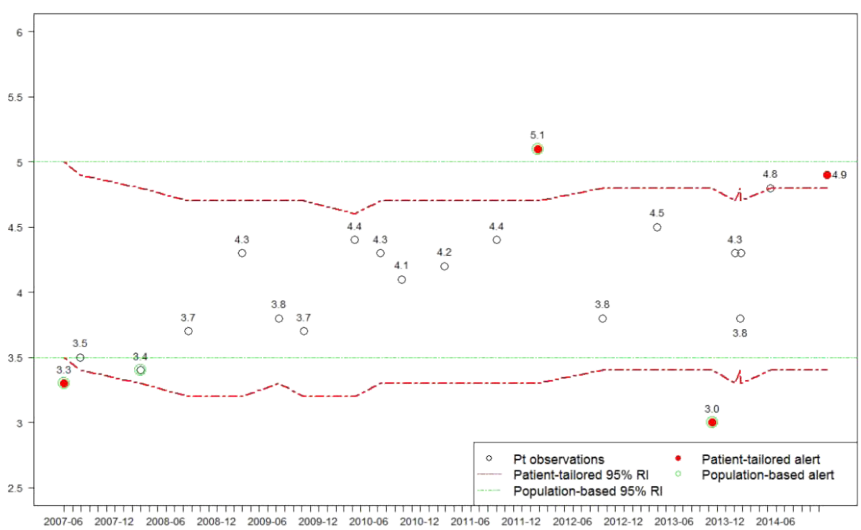


Figure 1 – Example of application of the patient-tailored and population-based RIs to a patient's potassium results time series to detect abnormal values.

Evaluation

Survey design

In order to test the clinical relevance of flagged values by the patient-tailored method as well as the standard method, we defined a gold standard as the GPs' clinical judgement, which we obtained through a survey, designed as follows. From the analysed testing dataset, we randomly selected 15 values for each possible combination of the two methods, i.e. values alerted by: 1) only the standard method; 2) only the patient-tailored method; 3) the standard AND patient-tailored methods at the same time (referred in the text as *combined method*); 4) none. From these values, we randomly included 2 per combination (for a total of 8) in a survey. GPs were asked to rate on a colour-based scale how abnormal each value was in the context of the specific patient:

- *Green* (normal value; i.e. no actions required);
- *Yellow* (probably abnormal; i.e. repeat in more than a week, do further test, change medication);
- *Red* (definitely abnormal; i.e. repeat urgently / hospital admission).

For each value GPs were provided with patient characteristics (i.e. age and gender), a graph showing the patient's previous potassium results, brief summary of past medical history (i.e. comorbidities and time since diagnosis), and medications from the last four months. In order to avoid any possible priming, no information about the standard and patient-tailored RIs were provided. Since values were randomly selected each time from the 15 values per combination included, it is noteworthy that: all surveys administered to participants were different between each other, while maintaining the proportion between the different combination of the two methods; each value could be assessed by more than one general practitioner.

The survey also contained three questions about respondent's working days (1-3 days, 4-5 days), years of experience (1-10 years, 10-20 years, >20 years), and opinion about abnormal test results alerts ("not enough", "about right", "too much")

Participants and survey setting

We administered the survey to a group of 43 GPs taking part in a 5 days Continuing Professional Development course for leadership development held in the city of Manchester (UK) in Octorber 2014.

Data analysis

We considered as clinically relevant, a value that was judged by GPs as at least probably abnormal (i.e. yellow and red in the coloured scale in the survey). In order to test performance, we calculated sensitivity and PPV for the standard, patient-tailored and combined methods by taking the prevalence of the different type of flagged values (i.e. flagged only by the standard method, only by the patient-tailored or by both methods at the same time) in the original dataset into account.

In order to assess variable importance, possible intra-assessor and intra-value correlation, a mixed-effects logistic regression was employed, using the flags by standard and patient-tailored methods as well as respondent characteristics as independent variables and the values clinical relevance as binary outcome.

All analyses were performed using the R software (http://www.r-project.org/).

Results

Data analysis

We extracted 1,411,757 unique potassium results from the SIR database, for a total of 151,681 patients. Mean age at first potassium record was 45.4 (17), female accounted for 49.8% and the mean follow-up time was 4.7 (4). The training dataset was composed of 500 patients and 4,144 potassium results. Of these, 470 (11.3%) values were flagged as abnormal by the standard method, 372 (9%) by the patient-tailored method and 301 (7.3%) by the combined one. The patient-tailored and combined methods registered a 20% (98 values) and 36% (169 values) reduction of the number of abnormal flagged values compared to the standard method.

Evaluation

Of the 43 GPs that received the survey, 19 completed it (response rate of 44%). Each value was assessed by a median of 3 GPs. Table 1 reports respondents' characteristics. The majority of GPs had more than 20 years of experience in general practice (63.2%) and worked 1-3 days in general practice. Furthermore, 42.1% thought that there are too many test results alerts in general practice.

Table 1 – Baseline characteristics of general practitioners.

Question	Reply	N (%)
Days per week in practice	1-3 days	10 (52.6%)
	4-5 days	9 (47.4%)
Years of experience	<10 years	2 (10.5%)
	10-20 years	5 (26.3%)
	>20 years	12 (63.2%)
Opinion about tests alerts in general practice	Not enough	4 (21.1%)
	About right	7 (36.8%)
	Too much	8 (42.1%)

Out of the 152 values assessed by GPs, 92 values were cosidered normal (green), 54 as probably abnormal (yellow), and 6 values as definitely abnormal (red). On average, each general practitioner identified 4.8 normal values (minimum: 1, maximum: 8), 2.8 probably abnormal values (minimum: 0; maximum: 7), and 0.3 definitely abnormal values (minimum: 0; maximum: 1).

Table 2 reports cross tabulation of values flagged as abnormal by all methods and whatever values were considered clinically relevant by GPs.

Table 2– Cross tabulation between standard, patient-tailored and combined methods and general practitioners judgement as clinically relevant.

		Clinically relevant		
		Neg	*Pos*	*Total*
Standard method	*Neg*	63	13	76
	Pos	29	47	76
	Total	92	60	152
		Clinically relevant		
		Neg	*Pos*	*Total*
Patient-tailored method	*Neg*	55	21	76
	Pos	37	39	76
	Total	92	60	152
		Clinically relevant		
		Neg	*Pos*	*Total*
Combined method	*Neg*	83	31	114
	Pos	9	29	38
	Total	92	60	152

Table 3 reports sensitivity and PPV of all three methods, calculated by taking the prevalence in the original dataset into account. The standard and combined methods had the best performance in terms of sensitiviy and PPV respectively.

Table 3– PPVs and sensitivities based on general practitioners judgements as clinically relevant.

Parameter	Standard method	Patient-tailored method	Combined method
Prevalence in testing dataset (n=4144)	11.3%	9%	7.3%
Prevalence in values assessed by GPs	50%	50%	25%
Sensitivity	0.51	0.41	0.38
PPV	0.66	0.67	0.76

Table 4 shows the adjusted ORs for the fixed effects in the mixed-effects logistic regression of values identified as clinically relevant by GPs. The estimated variance of the random effects for assessor and value were 1.5 (SD:1.2) and 0.4 (SD: 0.6) respectively. The only two significant variables were the flags by the standard and patient-tailored methods.

Table 4– Adjusted ORs for values identified as clinically relevant by general practitioners.(GP: General Practice)

Parameter	Adjusted OR [95% CI]
Standard method pos. vs neg.	24.5 [5.3,113.7]
Patient tailored method pos. vs. neg.	6.2 [2.0,19.1]
Weekly working days in GP: 4-5 days vs 1-3 days	2.2 [0.4,11.3]
Years of experience in GP: 10-20 years vs <10 years	3.5 [0.4,11.3]
Years of experience in GP: >20 years vs <10 years	6.0 [0.3,103.1]
Opinion about tests alerts in GP: not enough vs about right	0.5 [0.7,3.7]
Opinion about tests alerts in GP: too much vs about right	0.2 [0,1.3]

Discussion

This paper describes the development and preliminary evaluation of a method to produce dynamic, patient-tailored alerts for abnormal test results. The evaluation was focused on potassium results and clinical relevance was assessed via a survey administred to a group of experienced GPs in Manchester (UK).

Looking at performance of the standard and patient-tailored method, overall the standard method yielded a better trade-off between sensitivity and PPV. Although the dynamic, patient-tailored method achieved a reduction in the number of flagged abnormal values and a slightly better PPV, the standard method was significantly more sensitive.

The combined method showed poor sensitivity (similar in absolute numbers to the patient-tailored method). However, PPV was the best by far. This result suggests that combining population RI with patient-specific contextual information, improves the clinical relevance of flagged values. This finding is not unexpected: values flagged by the combined method are potassium results that are abnormal for the healthy population as well as in the context of a particular patient. The importance of the additional information provided by the patient-tailored method is also confirmed by the mixed-effects logistic regression modelling.

To decrease the number of not relevant test results alerts, some investigators have suggested alerting physicians only about those values that have a high deviance from the RI [16, 17]. We performed a sensitivity analysis by adopting a threshold-based approach with potassium RI of 2.5-6 mmol/l (previously proposed in [17]) that confirmed our main analysis. In detail, weighted sensitivity and PPV were 0.1 and 0.89 respectively. Although values outside the adopted RI had a very high deviance from the UK potassium RI, not all flagged values were considered clinically relevant and just a small proportion were considered definitely abnormal. This confirms the importance of the context and the bluntness of using fixed-thresholds when looking at individual patients.

Our main focus was on PPV, which is key to avoiding alert fatigue. However, we note that low sensitivities obtained by the patient-tailored and combined methods would lead to clinically relevant abnormal values being missed, with possible consequences for patient safety. It is noteworthy that in absolute numbers, even the sensitivity by the standard method (0.51) cannot be considered satisfactory and would lead to miss many values that are abnormal in the context of a specific patient but lay within the population RI. In addition to the lack of sensitivity, the patient-tailored and stardard methods have further drawbacks directly related to how they are defined. On one side, because of his adaptative nature, the patient-tailored method in presence of patients that are not stable but worsen over time (i.e. the value for a specific parameter keeps increasing or decreasing) would slowly adapt to the abnormal values and produce a biased patient-tailored RI. On the other side, the standard method, which uses fixed thresholds, would keep flagging tests results as abnormal when patients' values are stable but consistently outside the population RI. Improvements are needed in these regards for all methods.

Our analysis has several limitations. First, we focused on only one biochemical factor (potassium). Second, we carried out our evaluation within a relatively small panel of GPs in one geographical area. Third, the UK potassium RI adopted might have been slightly different than the one used in other laboratories that the clinicians may be used to; furthermore there might have been GPs that slavishly applied the abovementioned RI when replying to the survey.

Conclusion

This study demonstrates that the combined adoption of a patient-tailored method and a the standard threshold-based method for assessing potassium levels can improve the PPV of results flagged as abnormal. This could be particularly important to prioritise alerts by making the values flagged by both methods more prominent.

We plan to extend our experiments to a wider panel of laboratory tests (e.g. creatinine, eGFR, calcium, heamoglobin) and to a larger number of GPs, as well as investigating alternative statistical approaches (including Bayesian inference). We also plan to relate the alerting performance to adverse health outcomes.

This study represents a first step towards a next generation of context-aware alerting systems that in future may enhance patient safety.

Acknowledgements

Funded by the National Institute for Health Research Greater Manchester Primary Care Patient Safety Translational Research Centre (NIHR GM PSTRC) and the MRC Health eResearch Centre, Farr Institute, UK (MR/K006665/1). The

views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

[1] Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. J. Gen. Intern. Med. [Internet]. 2012 Oct [cited 2014 Oct 21];27(10):1334–48. Available from: http://link.springer.com/article/10.1007%2Fs11606-011-1949-5.

[2] Elder NC, Dovey SM. Classification of medical errors and preventable adverse events in primary care: a synthesis of the literature. J. Fam. Pract. [Internet]. 2002 Nov;51(11):927–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12485545.

[3] McKay J, Bradley N, Lough M, Bowie P. A review of significant events analysed in general practice: implications for the quality and safety of patient care. BMC Fam. Pract. [Internet]. 2009 Jan;10(1):61. Available from: http://www.biomedcentral.com/1471-2296/10/61

[4] Sittig DF, Singh H. Electronic Health Records and National Patient-Safety Goals. N. Engl. J. Med. [Internet]. Massachusetts Medical Society; 2012 Nov 7;367(19):1854–60. Available from: http://dx.doi.org/10.1056/NEJMsb1205420.

[5] Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. Arch. Intern. Med. [Internet]. American Medical Association; 2004 Nov 8 [cited 2014 Oct 16];164(20):2223–8. Available from: http://archinte.jamanetwork.com/article.aspx?articleid=217621.

[6] Murphy DR, Reis B, Sittig DF, Singh H. Notifications received by primary care practitioners in electronic health records: a taxonomy and time analysis. Am. J. Med. [Internet]. 2012 Feb;125(2):209.e1–7. Available from: http://www.sciencedirect.com/science/article/pii/S0002934311006474.

[7] Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern. Med. [Internet]. 2013 Apr 22;173(8):702–4. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3822526&tool=pmcentrez&rendertype=abstract.

[8] Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Smith MW, Murphy DR, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. J. Am. Med. Inform. Assoc. [Internet]. 2013;20(4):727–35. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3721157&tool=pmcentrez&rendertype=abstract.

[9] Sijs IH van der (Heleen). Drug safety alerting in computerized physician order entry: Unraveling and counteracting alert fatigue. 2009 Sep 2; Available from: http://www.webcitation.org/6NYq64xRe.

[10] Ceriotti F. Prerequisites for use of common reference intervals. Clin. Biochem. Rev. [Internet]. 2007 Aug [cited 2014 Oct 20];28(3):115–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1994109&tool=pmcentrez&rendertype=abstract.

[11] Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? Am. J. Clin. Pathol. [Internet]. 2010 Feb 1;133(2):180–6. Available from: http://ajcp.ascpjournals.org/content/133/2/180.long.

[12] Baron JM, Dighe AS, Arnaout R, Balis UJ, Black-Schaffer WS, Carter AB, et al. The 2013 symposium on pathology data integration and clinical decision support and the current state of field. J. Pathol. Inform. [Internet]. 2014 Jan;5:2. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3952400&tool=pmcentrez&rendertype=abstract.

[13] Galecki A, Burzykowski T. Linear Mixed-Effects Models Using R A Step-by-Step Approach. Casella G, Fienberg SE, Olkin I, editors. New York, New York, USA: Springer New York; 2012.

[14] NHS England. Read Codes [Internet]. Available from: http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes.

[15] Reference range (potassium) - General Practice Notebook [Internet]. Available from:http://www.gpnotebook.co.uk/simplepage.cfm?ID=584056899.

[16] Salinas M, López-Garrigós M, Gutiérrez M, Lugo J, Flors L, Leiva-Salinas C. Should we customise critical value procedure according to patient origin and laboratory turnaround time? J. Clin. Pathol. [Internet]. 2013 Apr 1;66(4):269–72. Available from: http://jcp.bmj.com/content/66/4/269.

[17] Mathew G, Kho A, Dexter P, Bloodworth N, Fantz C, Spell N, et al. Concept and development of a discharge alert filter for abnormal laboratory values coupled with computerized provider order entry: a tool for quality improvement and hospital risk management. J. Patient Saf. [Internet]. 2012 Jun;8(2):69–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22543363.

Address for correspondence

Corresponding author: Paolo Fraccaro, MEng.

Address: Vaughan House, The University of Manchester, Portsmouth St, Manchester M13 9GB

Email address: paolo.fraccaro@postgrad.manchester.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-706

Automated Detection of Postoperative Surgical Site Infections Using Supervised Methods with Electronic Health Record Data

Zhen Hu[a], Gyorgy J. Simon[a], Elliot G. Arsoniadis [a,b], Yan Wang[a], Mary R. Kwaan[b], Genevieve B. Melton[a, b]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[b] *Department of Surgery, University of Minnesota, Minneapolis, MN, USA*

Abstract

The National Surgical Quality Improvement Project (NSQIP) is widely recognized as "the best in the nation" surgical quality improvement resource in the United States. In particular, it rigorously defines postoperative morbidity outcomes, including surgical adverse events occurring within 30 days of surgery. Due to its manual yet expensive construction process, the NSQIP registry is of exceptionally high quality, but its high cost remains a significant bottleneck to NSQIP's wider dissemination. In this work, we propose an automated surgical adverse events detection tool, aimed at accelerating the process of extracting postoperative outcomes from medical charts. As a prototype system, we combined local EHR data with the NSQIP gold standard outcomes and developed machine learned models to retrospectively detect Surgical Site Infections (SSI), a particular family of adverse events that NSQIP extracts. The built models have high specificity (from 0.788 to 0.988) as well as very high negative predictive values (>0.98), reliably eliminating the vast majority of patients without SSI, thereby significantly reducing the NSQIP extractors' burden.

Keywords:

Electronic Health Records, National Surgical Quality Improvement Project, Supervised Learning, Surgical Site Infection

Introduction

The American College of Surgeons (ACS) National Surgical Quality Improvement Project (NSQIP) is widely recognized as "the best in the nation" surgical quality improvement resource in the United States [1]. NSQIP helps member hospitals to track outcomes associated with surgical patients, by collecting data on over 150 variables, including preoperative characteristics, intraoperative factors, and postoperative morbidity occurrences. In particular, postoperative morbidity outcomes are rigorously defined surgical adverse events occurring within 30 days of surgery, such as surgical site infection (SSI), urinary tract infection (UTI), and acute renal failure (ARF). NSQIP uses collected data elements to calculate relative performance regarding postoperative morbidity and mortality and to compare each member hospital's performance with benchmarking, which is risk stratified, including providing an observed to expected (O/E ratio) for every surgical adverse event [2]. With this feedback, member hospitals are able to focus attention and resources to areas of opportunity for improving the care of patients, which may also result in achieving reduced length of stay and readmission rates [3].

Unfortunately, less than 20% of hospitals in the United States currently participate in NSQIP, in large part due to its associated costs to implement. In addition to the participation fee, hospitals must employ a formally trained surgical clinical reviewer (SCR). An SCR ensures the reliability of clinical data abstraction, selects operation cases following NSQIP inclusion criteria, manually reviews and extracts data elements, and documents surgical postoperative occurrence outcomes. This manual yet expensive approach leads to high-quality clinical data, but the associated cost remains a significant bottleneck to NSQIP's wider dissemination.

An SSI is an infection occurring after surgery in the part of the body where surgery took place. While most surgical patients do not experience an SSI [4], SSIs are very expensive and morbid. According to the depth and severity of infection, SSIs are categorized into superficial, deep, and organ/space. Definitions for SSIs have been standardized by the Centers for Disease Control and Prevention (CDC) and are used by NSQIP SCR to identify and document each SSI category [5-6].

Previous work has explored risk factors associated with SSI, but few studies have focused on the detection of SSI. Most papers examining detection have relied heavily on administrative data or claims databases (such as age, gender, principal diagnosis, and billing information about medications and procedures) [7-8]. Since EHR data contains more detailed and richer clinical data (e.g. vital signs, lab results, and social history), compared with claims data it would provide additional significant indicators and signals to SSI and thus enhance the detection performance. In addition, most studies are procedure-specific, only processing SSIs following certain types of operation, such as hip and knee arthroplasty [9-10], instead of the current approach which is broadly inclusive of different types of surgery. To help reduce the labor and cost in reviewing patient records for postoperative surgical occurrences, we hypothesized that we could leverage both electronic health record (EHR) data and historic NSQIP registry data to develop and validate an automated approach with supervised machine learning algorithms to detect NSQIP occurrence outcomes. In particular, we focused on the postoperative SSI occurrences to develop a classifier of three SSI categories (superficial, deep, and organ/space) and the total SSI, and to reduce the SCR's burden by eliminating the vast majority of patients associated with surgeries that did not result in SSI.

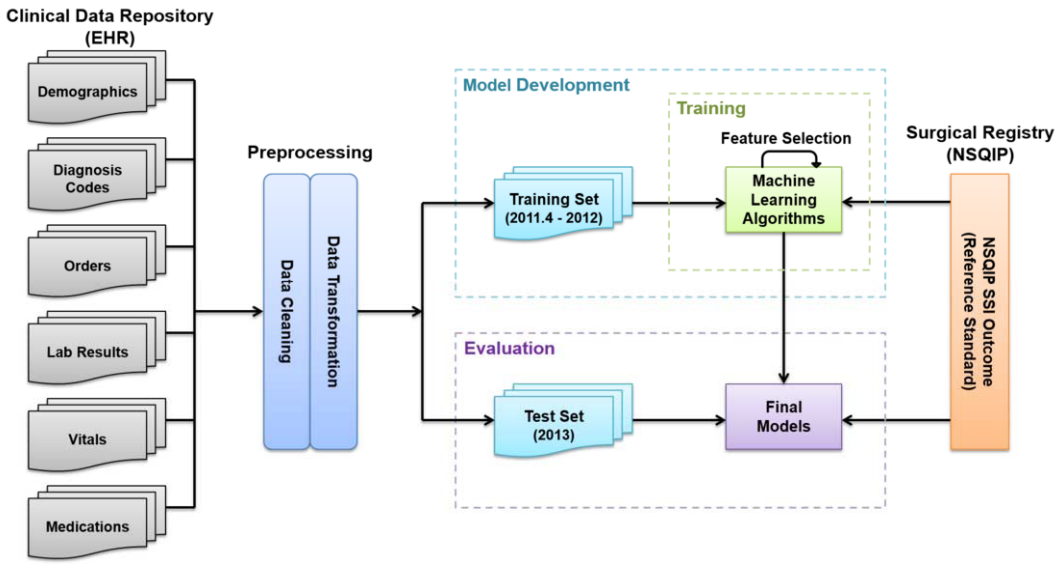


Figure 1 - Overview of Materials and Methods

Materials and Methods

Our overall methodological approach for this study included four steps as outlined in Figure 1: (1) identification of the patient cohort and associated patient EHR data, (2) data preprocessing, (3) iterative supervised learning model development, and (4) evaluation of the final models using gold standard outcome data from the NSQIP registry. Institutional review board approval was obtained and informed consent waived for this minimal risk study.

Data Collection and Patient Cohort Identification

The University of Minnesota Academic Health Center Information Exchange platform includes access to the clinical data repository (CDR) which contains University of Minnesota Medical Center (UMMC) clinical data. UMMC has been a member of NSQIP since 2007 and has used the inpatient instance of Epic since April 2011. CDR and NSQIP registry, two different data sources, were linked through the patient medical record number and the date of surgery. Subjects with no records in the EHR were eliminated. The patient cohort was divided into two datasets: data of patients with surgery from April 2011 to the end of 2012 (model development set) and data of patients with surgery in 2013 (evaluation set). The former dataset was used as the training set for model development. The evaluation dataset was held out fully for the overall evaluation of the developed models. Table 1 describes the detailed demographic information. From April 1, 2011, through December 31, 2013, a total of 6258 procedures with 405 SSIs were collected. The period of April 2011 to the end of 2012 comprised 3996 procedures and 278 SSIs (6.95% rate). About 79% procedures were patients no more than 65 years old, and 21% were patients more than 65 years old. Approximately 83.8% were white, 8.6% were black, and 7.6% were other race/ethnicity and unknown. The year of 2013 comprised 2262 procedures and 127 SSIs (5.6% rate), with similar patient characteristics, as shown in Table 1.

The clinical data utilized included six data types: demographics, diagnosis codes, orders, lab results, vital signs, and medications. Demographics included each patient's gender, race, and age at the time of surgery. Diagnosis codes consisted of related ICD-9 codes generated during the encounter and hospital stay at the time of surgery from coding, as well as the diagnoses from the past medical history and problem list. Orders related with SSI diagnosis and treatment were also gathered from the EHR, including imaging orders, infectious disease consultation orders, and procedures with incision and drainage. The most recent lab values and vitals results before surgery and those during the postoperative 30-day window (since surgical adverse events defined as occurring within 30 days after surgery) were collected. Medications utilized for this analysis included antibiotics from the third day after surgery onwards.

Another two important data measures included were American Society of Anesthesiologists (ASA) physical status classification and surgical wound classification. ASA classification from 1 to 6 indicates a patient's status from normal healthy to declared brain-dead; the surgical wound classification is used for postoperatively grading of the extent of microbial contamination, indicating the chance a patient will develop an infection at the surgical site. We dichotomized the wound classification as the bottom two (no or mild disturbance) versus the remaining levels (significant disturbance).

Data Preprocessing

EHR data of interest were collected, cleaned, and analyzed next. Identifying and removing outliers, and correcting inconsistent data were the very first tasks of data preprocessing. How to transform clinical data into meaningful features was our main interest. Most clinical data, such as lab test results and vitals, tended to be longitudinal with repeated measures. Traditional methods to summarize those variables by calculating the moments (mean and standard deviation) or extremes tended not to be sufficient to describe the temporal behavior of such variables. To better summarize individual tests, we explored other features like the change of values

Table 1 - Characteristics of patients and cases of surgical site infection (SSI) from cohorts

Characteristic	4/2011-2012					2013				
	No. of procedure	No. of SSI	Superficial SSI	Deep SSI	Organ/Space SSI	No. of procedure	No. of SSI	Superficial SSI	Deep SSI	Organ/Space SSI
Total	3996	278	140	52	86	2262	127	47	35	45
Encounter type										
Inpatient	3018	265	147	47	83	1352	108	36	32	40
Outpatient	978	13	5	5	3	910	19	11	3	5
Age group										
< 65	3160	210	112	41	67	1569	90	28	28	34
≧65	836	68	38	11	19	693	37	19	7	11
Gender										
Male	1530	122	56	19	47	974	56	21	15	20
Female	2466	156	84	33	39	1288	71	26	20	25
Race										
White	3366	226	115	45	66	1881	110	38	31	41
Black	269	23	10	5	8	142	8	4	1	3
Other	361	29	15	2	9	239	9	5	3	1

during an "elevating period". An elevating period is a time period during which the measurement in question is near-monotonously increasing from a low level (trough point) to a high level (peak point). For patients with SSI, some lab results, like serum glucose (GLC), platelet count (PLT), and white blood cells (WBC), have significant increases in the measurement from the third day after operation.

As shown in Figure 2, GLC increased in three time periods: (I) day 3~7, GLC increased from 116 to 128; (II) day 7~9, from 104 to 140; and (III) day 15~28, from 87 to 148. Such elevation may indicate the onset of SSI. To capture the elevating period, a feature defined as the postoperative increase from a trough to its nearest peak was included in our tentative model. In the case of multiple elevating periods, the feature was computed by using the period with the highest peak. For measures where low values could indicate SSI, a "descending period" can be defined analogously.

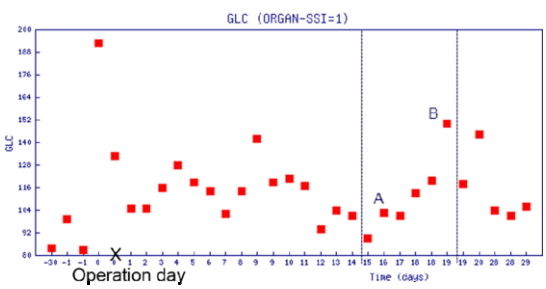


Figure 2 - GLC values within 30 days before and after surgery

Figure 3 depicts the flowchart of the algorithm to compute this feature. The algorithm first searches for the maximum value (p_m) from all results at least two days after the operation ($\{p_i, i=0, \ldots,n\}$), (e.g., in Figure 2, point B is the maximum GLC value, which was measured nineteen days after the operation). Then the algorithm proceeds by searching for the trough point backward from point B. The algorithm is robust in its filtering of the abnormal point that temporarily breaks the rule of monotone. For example, in Figure 2, the elevating period is

from day 15 to 19, however, there is an abnormal point A which breaks the monotone increasing trend between day 15 and 17; to overcome the problem and identify the real trough, the algorithm further compares day 15 and day 17 in order to determine whether the criterion of monotone increasing is satisfied.

For other data like antibiotic use and specific orders, we created binary variables to indicate whether a relevant element was observed. For example, a value of 1 for Interventional Radiology signifies that an abscess drainage order was placed for a patient; while a value of 0 signifies that no such test was ordered.

Model Development

To build our SSI detection model, we utilized multivariate logistic regression models. We constructed one model for total SSI and one model for each of the three SSI subtypes. Binary variables were entered as dummy indicator variables and continuous variables were entered unmodified. We used stepwise construction to select significant features and Akaike Information Criterion (AIC) for model selection.

Evaluation

In assessing detection of surgical adverse event outcomes like SSI, since these events are relatively rare, overall detection accuracy percentage is not an optimal criterion for evaluating model validity. Instead, we report specificity, as well as the the area under the curve (AUC), in evaluation of our automated detection system. Our aim was to maximize the specificity under the constraint that the negative predictive value remains above 98%. This aim is reflective of our original expectation of actual use of the detection models: to assist a NSQIP chart extractor to eliminate patients who clearly did not suffer the adverse event and then to accelerate the process of data abstraction from clinical charts.

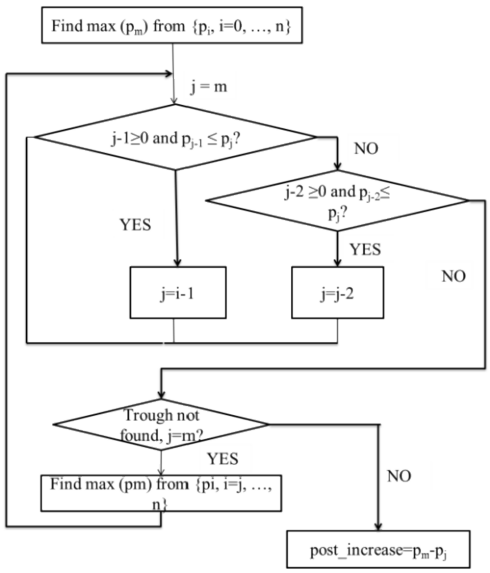


Figure 3 - Finding the postoperative increase in GLC

Results

Significant Variables Selected

Tables 2 through 5 show the results for the multivariate detection models for the three kinds of SSI and the total SSI, selected by AIC. The two most common variables included were diagnosis codes (the ICD-9 codes of SSI is 998.xx) and antibiotic use. Superficial SSI occurs just at the skin incision and thus relatively easily diagnosed. Therefore, imaging diagnostic orders tend to be unnecessary. Infection is sometimes diagnosed with microbiology cultures, however, frequently this diagnosis is based on the physical examination only. Actually only cultures ordered or not is a signal of SSI. According to table 3 and table 4, we can find that abscess culture, fluid culture and wound culture are significant factors for detecting deep and organ/space SSI. Since these two kinds of SSIs occur deep within or under the wound, imaging orders for both diagnosis and treatment are frequently required.

We also found the postoperative elevating period of GLC, for superficial and PLT for organ/space, to be indicative of clinical suspicion. Clinically these lab values can be altered in the setting of infection. For a unit increase in postoperative increase of GLC, we expect to see approximately a 0.0112 increase in log-odds of superficial SSI. Similarly, for a unit postoperative increase of PLT, approximately a 0.0115 increase in the log-odds of organ/space SSI is expected.

Table 2 - Significant indicators for detecting superficial SSI

Significant variables	Estimate	P-value
Diagnosis codes	2.1126	<0.0001
Wound culture ordered	2.1941	<0.0001
Antibiotic use	1.1321	<0.0001
Encounter type (inpatient)	1.6007	0.0010
ASA Classification (significant disturbance)	0.4342	0.0058
Abscess culture ordered	1.5020	0.0050
Postoperative increase of GLC	0.0112	0.0687

Table 3 - Significant indicators for detecting deep SSI

Significant Variables	Estimate	p-value
Diagnosis codes	3.1959	<0.0001
Antibiotic Use	2.2276	<0.0001
Abscess culture ordered	1.2880	0.0868
Gram stain ordered	0.8040	0.0427
Imaging treatment ordered	1.5445	0.1107
Imaging diagnosis ordered	0.6254	0.0981
Tissue culture ordered	1.6516	0.1010

Table 4 - Significant indicators for detecting organ/space SSI

Significant Variables	Estimate	p-value
Imaging treatment	1.3999	<0.0001
Imaging diagnosis	1.2090	<0.0001
Antibiotic Use	1.1662	<0.0001
Abscess culture ordered	2.3041	<0.0001
Fluid culture ordered	1.4204	0.0003
Preoperative PLT	0.00332	0.0135
Drainage culture ordered	1.3760	0.0711
Diagnosis code	0.8259	0.0667
Postoperative increase of PLT	0.0115	0.0606

Table 5 - Significant indicators for detecting total SSI

Significant Variables	Estimate	p-value
Diagnosis codes	5.3940	<0.0001
Antibiotic use	1.3672	<0.0001
Abscess culture ordered	3.2565	<0.0001
Wound culture ordered	2.2926	<0.0001
Imaging diagnosis ordered	0.8741	<0.0001
Fluid culture ordered	1.2909	<0.0001
Encounter type (inpatient)	1.0185	0.0037
ASA Classification (significant disturbance)	0.4258	0.0031
Preoperative PLT	0.00214	0.0440
Post maximum pain	0.0775	0.0957

Model Performance

Four detection models exhibited excellent specificity to eliminate the majority of non-SSI patients, which greatly accelerate the process of extracting postoperative SSI occurrences. Table 6 presents the negative predictive value (NPV) for each of the SSI identification models. The highest specificity 0.988 was for detecting deep SSI at NPV equals to 0.99, and the lowest 0.787 was for detecting total SSI at NPV equals to 0.99. AUC values for the four models were 0.820, 0.898, 0.886 and 0.896, respectively.

Table 6 - Negative predictive value and specificity for four SSI models

	NPV	Specificity
Superficial SSI	0.980	1.000
	0.985	0.987
	0.990	0.900
Deep SSI	0.980	1.000
	0.985	1.000
	0.990	0.988
Organ/space SSI	0.980	1.000
	0.985	0.999
	0.990	0.974
Total SSI	0.980	0.935
	0.985	0.888
	0.990	0.787

Discussion

The current research is a pilot study to examine the feasibility of automatically detecting postoperative SSI occurrences based on EHR data. The aim of this study is to assist a NSQIP SCR to eliminate patients who clearly did not suffer the adverse event. Therefore, a very high NPV is desired, which could assist in the reliable identification of patients without postoperative SSI. From the modeling results, we can see that all four models perform very well (with specificity ranging from 0.788 to 0.988) in eliminating the majority of patients without SSI based on the NPV equals to 0.99. Considering the nature of NSQIP SCR's work, SCRs still need to review all clinical charts, even if the positive predictive value for a patient is 0.9 or higher, since they need to extract the clinical characteristics of patients with SSI. Therefore, achieving high NPV, and thus allowing SCRs to eliminate patients, rather than achieving a high positive predictive value, is the main focus of this research.

Among selected potential indicators, a few of them were found to be quite significant with very small p-values. Only the indicators that had p-value less than 0.0001 were employed in the logistic regression modeling, however, this did not improve the detection performance. Other modeling methods, like Random Forest and Support Vector Machine, were employed; however, logistic regression models were found to outperform these methods for detection of all types of postoperative SSI events.

The current study was limited by the fact that it was conducted with only complete cases over three years. This may have limited our ability to fully refine and optimize the automated detection model. In the future, more procedures will be included, and the treatment of missing data will be studied.

Large quantities of meaningful information are stored at the clinical notes, such as imaging reports and culture results, which we did not utilize in this study. For example, a positive abscess culture result could be recorded as "On day 2, isolated in broth only: Bacteroides fragilis group". However, we merely considered whether the diagnostic and therapeutic imaging orders or cultures were placed, we did not use the actual results. Natural language processing (NLP) has played an important role in detecting adverse events [11-12]. In our future research, we will apply NLP techniques to extract additional important information from clinical notes.

Conclusion

In this study, to accelerate the process of extracting postoperative SSI outcomes from medical charts and reduce the workload of NSIQP SCR, an automated postoperative SSI detection model based on supervised learning was proposed and validated. The models exhibited good performance, they reduced the SCR's burden by reliably eliminating the vast majority of patients with no SSI. The significant factors of detecting SSI identified by our models are in line with clinical knowledge. In addition, some useful patterns, (e.g. postoperative increase of PLT and GLC), were extracted from the longitudinal lab results.

Acknowledgments

This research was supported by the University of Minnesota Academic Health Center Faculty Development Award (GS, GM) and the American Surgical Association Foundation (GM). The authors thank Fairview Health Services for support of this research.

References

[1] Fetter RB, Shin Y, Freeman JL, Averill RF, and Thompson JD. Casemix definition by diagnosis-related groups. Med Care 1980: 18 (2 Suppl): 1-53.

[2] Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk-adjustment for comparing hospital quality with surgery: How many variables are needed? J Am Coll Surg. 2010 April; 210(4):503-508.

[3] Guillamondegui OD, Gunter OL, Hines L, Martin BJ, Gibson W, Chris Clarke P, Cecil WT, Cofer JB. Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to Improve Surgical Outcomes. J Am Coll Srug.2012 Apr;2014(4):709-14.

[4] Rutala WA, Weber DJ. Cleaning, disinfection, and sterilization in healthcare facilities: What clinics need to know. Clin Infect Dis. 2004 Sep 1; 39(5):702-9.

[5] http://www.cdc.gov/nhsn/acute-care-hospital/ssi/

[6] http://site.acsnsqip.org/wp-content/uploads/2012/03/ACS-NSQIP-Participant-User-Data-File-UserGuide_06.pdf.

[7] Mu Y, Edwards JR, Horan TC, Berrios-Torres SI, Fridkin SK. Improving risk-adjusted measures of surgical site infection for the national healthcare safety network. Infect Control Hosp Epidemiol. 2011 Oct; 32(10):970-86.

[8] Levine PJ, Elman MR, Kullar R, Townes JM, Bearden DT, Vilches-Tran R, McClellan I, McGregor JC. Use of electronic health record data to identify skin and soft tissue infections in primary care settings: a validation study. BMC Infectious Disease. 2013 Apr 10; 13:171.

[9] Warren DK, Nickel KB, Wallace AE, Mines D, Frasesr VJ, Olsen MA. Can additional information be obtained from claims data to support surgical site infection diagnosis codes? Infect Control Hosp Epidemiol. 2014 Oct; 35 Suppl 3: S124-32.

[10] Price CS, Satitz LA. Improving the measurement of surgical site infection risk stratification/outcome detection. Final report (prepared by Denve health and its partners under contract No. 290-2006-00-20). AHRQ publication No. 12-0046-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2012.

[11] Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc. 2005 Jul-Aug; 12(4): 448-57.

[12] Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. J Am med Inform Assoc. 2003 Mar-Apr; 10(2):115-28.

Address for correspondence

Genevieve B. Melton, MD., M.A. Faculty Fellow, Institute for Health Informatics, and Associate Professor, Department of Surgery, University of Minnesota, MMC 450, 420 Delaware Street SE, Minneapolis, MN, 55455, USA. Email: gmelton@umn.edu.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-711

Mining Temporal and Data Constraints Associated with Outcomes for Care Pathways

Xiang Li[a], Haifeng Liu[a], Jing Mei[a], Yiqin Yu[a], Guotong Xie[a]

[a] *IBM Research, Beijing, China*

Abstract

A care/clinical pathway defines a standardized care process for a specific patient group, which consists of clinical goals, activities, data attributes, and constraints describing temporal dependencies and data preconditions of the activities. The constraints, which are the key elements to represent the best practices, are difficult to define due to the variations in different regions and populations. In this paper, we propose an approach to discover temporal and data constraints that are correlated with clinical outcomes for care pathways. For each activity of interest, we extract a set of associated event-condition-action (ECA) rules from electronic medical records (EMR) to represent the temporal and data preconditions of the activity, by using our modified association rule mining algorithm. Then the best ECA rule that is significantly more likely to lead to a positive outcome is translated into the constraint on the activity. The approach has been applied to real-world EMR, and discovered meaningful constraints for different groups of type 2 diabetes patients, which can be used to provide decision support during individual patient care.

Keywords:

Clinical Pathways; Data Mining; Association Learning; Electronic Medical Record.

Introduction

A care/clinical pathway defines a standardized care process for a well-defined patient group, aimed at improving patient outcomes and promoting patient safety [1]. Generally, a care pathway consists of clinical goals (e.g., controlling glucose of diabetes patients), activities such as laboratory tests and medications, data attributes such as problem histories and test items, as well as constraints on the activities. The constraints, which describe temporal dependencies (e.g., check renal function (RFT) before using biguanides (BG)) and data preconditions (e.g., BG can be used only if the patient does not have ketoacidosis (KA) and creatinine (CR) ≤ 2.0 mg/dL) of the activities, play the most important role in achieving the clinical goals and reducing medical errors.

Real-world care processes, however, are dynamic, complex and ad-hoc [2]. For a certain disease, the care practices may vary in different regions, different periods and different patient populations. Therefore, if only guided by general medical knowledge such as clinical guidelines, it is usually difficult to define the pathway constraints that can represent the best practices for a specific patient group. However, because real clinical activities and patient outcomes are tracked in electronic medical records (EMR), a potential improvement for care pathway modeling is to discover the constraints from EMR using data mining and process mining techniques.

It is a challenging problem, however, to discover the complex constraints for a pathway. Firstly, both temporal dependencies and data preconditions should be extracted to build the care pathway. These temporal and data preconditions are usually associated with each other and cannot be mined independently. Secondly, to summarize the best practices for a specific patient group, the discovered constraints should be correlated with positive clinical outcomes. Previous works on clinical process mining [2-5] mainly focused on mining temporal dependencies of activities, and did not consider data conditions. Moreover, most of these methods [2-4] focused on the extraction of frequently satisfied dependencies. These frequent practices are not always the best practices. Although an existing solution [5] can highlight the mined temporal dependencies that have high correlation with patient outcomes, it can neither distinguish positive dependencies from negative ones, nor discover the data conditions associated with the outcomes. Several process mining methods for general purposes [6,7] can discover data preconditions that discriminate between satisfaction and violation of a temporal constraint, but these methods cannot obtain constraints that are correlated with outcomes.

In this paper, we address these issues by proposing an approach to discover constraints that are associated with positive clinical outcomes for care pathways. The constraints are defined using event-condition-action (ECA) rules, where temporal dependencies and data preconditions of an activity can be represented as events and conditions respectively. We discover the ECA rules from EMR by providing a modified association rule mining algorithm, which uses the metrics of support and valid rate to measure the frequency of a rule, as well as the relative risk and Fisher's exact test to quantify its correlation with a clinical outcome. For each activity of interest, the best ECA rule, which is significantly more likely to cause a positive outcome, is translated into the constraint of the activity. The proposed approach has been applied to real-world EMR to extract constraints for different groups of type 2 diabetes patients. The discovered constraints are reasonable and meaningful, which can be integrated into a care pathway management system to help clinicians refine the pathways and provide decision support during individual patient care.

Methods

Our method takes as input, a set of patient traces from the EMR and a draft care pathway, and the output is a complete care pathway with discovered constraints that are significantly associated with positive outcomes. The method includes two main steps. Firstly, we construct an activity dataset from the patient traces to represent the execution situation of the care pathway during each activity of interest. Secondly, a set of ECA-based constraints are mined from the activity dataset.

Materials

A variety of representations have been proposed for care pathway modeling, such as computer-interpretable guideline

languages [8], business process models [5], and the case management modeling and notation (CMMN) standard [9]. Generally, the key elements, such as stages, goals, activities, data attributes, and constraints on the activities are normally defined in a care pathway model. Actually, using our previous approach [10], every patient trace can be aligned with a multistage care pathway, where each record is mapped onto a corresponding stage according to a hidden Markov model. In this work, we focus on the constraint discovery within each stage of the pathway. Formally, a stage of a pathway model can be represented as $M = (g, A, B, \Phi)$, where g is the clinical goal (e.g., Glycated hemoglobin (A1c) < 7.0% and Fasting plasma glucose (FPG) ≥ 3.9 mmol/L), A is a set of clinical activities, B is a set of data attributes and Φ is a set of constraints on the activities of interest $A_I \subseteq A$. A constraint $\varphi \in \Phi$ can be defined in the form of an ECA rule:

on *events*(A) **if** *conditions*(B) **do** *action*(a_I),

where the event (on) part represents the temporal dependency of the activity $a_I \in A_I$ on other activities in A, and the condition (if) part represents the data preconditions of the activity a_I. We treat each event, condition or action as a rule item $c = (f, \omega_f, x_f)$, where $f \in A \cup B$ is the feature to represent an activity or an attribute, $\omega_f \in \Omega_f$ is the possible operator on f (e.g., =, ≥, ≤), and x_f is the reference value that can be boolean, categorical or numeric depending on the feature type. A ECA rule contains a set of rule items $C = C_A \cup C_B \cup C_I$, where C_A, C_B and C_I are respectively the item sets in the event, condition and action parts of an ECA rule. For instance, the above example about biguanides can be formally represented as:

on *Renal_function_test*.completed = true
if *Ketoacidosis* = *false* ∧ *Creatinine* ≤ 2.0
do *Biguanides*.enabled = true.

Figure 1 shows the graphical representation of this constraint in a CMMN-based care pathway model [9,11].

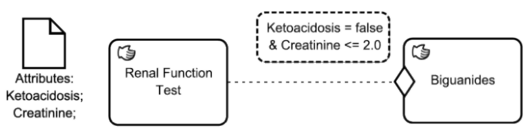


Figure 1 – A fragment of an example care pathway model

In EMR, a patient record can be formally represented as $r = (p, e, v, t)$, where $p \in P$ is the patient identifier from a cohort P, $e \in E$ is the code from a standard terminology to represent a care event such as a diagnosis, a medication or a lab test, v is the value (e.g., the value of a test item), and t is the occurring time of the event. For a specific patient p, his/her records can be sorted by t to generate a patient trace $\sigma = \langle r_1, r_2, \ldots, r_k \rangle$. Figure 2 illustrates an example patient trace, where "CR (1.0)" is a record whose code e represents CR and $v = 1.0$.

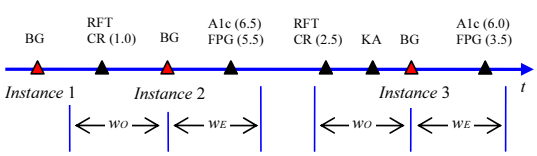


Figure 2 – An example patient trace

In a draft pathway model, the goal g, activities A and attributes B are pre-defined by users, while the constraint set Φ is empty. For constraint mining, a many-to-one terminology mapping Δ_A: $E_A \rightarrow A$ between a subset of event codes $E_A \subset E$ and the activities A in the pathway should be pre-defined, as well as a mapping Δ_B: $E_B \rightarrow B$ between a subset of codes $E_B \subset E$ and the attributes B. With these mappings, every relevant patient record can be mapped onto an activity and/or an attribute. Then the objective of our approach is to discover a set of constraints Φ for the activities of interest A_I, from the given set of patient traces Σ. The mined constraints should be frequently fulfilled in Σ and associated with achievement of the goal g.

Activity Dataset Construction

To discover the desired constraints, we first derive an activity dataset D from patient traces Σ, where each instance $d \in D$ can represent the execution situation of the care pathway at the time point of a patient record r. An instance $d = (a_d, \boldsymbol{x}_A, \boldsymbol{x}_B, y)$ contains the current activated action a_d for an activity of interest, the event feature vector $\boldsymbol{x}_A$, the condition feature vector $\boldsymbol{x}_B$ and the outcome y after the action. Each event feature x_a in $\boldsymbol{x}_A$ is a boolean attribute to represent whether the corresponding activity $a \in A$ was executed within an observation time window w_O (e.g., 1 year) before the current action. Each condition feature x_b in $\boldsymbol{x}_B$ contains the most recent value of an attribute $b \in B$ within the observation window w_O. The outcome y is a boolean label to represent the achievement result of the goal g within an evaluation window w_E (e.g., 3 months) after the action a_d.

Table 1 shows an example activity dataset where the three instances are derived from the patient trace shown in Figure 2. In this example, BG is taken as the activity of interest. For each record that can be mapped onto BG (highlighted in Figure 2), we create a new instance to record at the time of that action, whether renal function test (RFT) and other activities had been executed within w_O, whether ketoacidosis (KA) and other problems had occurred, the most recent values of creatinine (CR) and other test items within w_O, and the outcome (the truth value of A1c < 7.0 ∧ FPG ≥ 3.9) after the action within w_E. Note that the condition features and the outcome can have null values, when they cannot be validly evaluated within the time windows.

Table 1 – An example activity dataset derived from Figure 2

Ins.	Action	Event features			Condition features			Outcome
	a_d	x_{a1} (RFT)		…	x_{b1} (KA)	x_{b2} (CR)	…	y
1	BG	false		…	false	null	…	null
2	BG	true		…	false	1.0	…	true
3	BG	true		…	true	2.5	…	false
…	…	…			…			…

Constraint Mining

Based on the activity dataset D, we then discover the desired constraints Φ. For each activity of interest $a_I \in A_I$, the objective is to mine an ECA rule $C \rightarrow O$ that is composed of a set of associated rule items $C = C_A \cup C_B \cup C_I$ and prone to result in a positive outcome O. Existing association rule mining methods [12, 13] can find a rule that satisfies *support* and *confidence* constraints based on the *Apriori* algorithm. Let $D(C) = \{d \mid d \in D, C(d) = \text{true}\}$ denote the set of instances satisfying all rule items of C, and $|D(C)|$ is the number of the instances, then the support of the rule is the proportion of instances satisfying the rule in the instances where the action a_d equals a_I:

$$Sup(C) = |D(C)| / |D(C_I)|,$$

where $D(C_I) = \{d \mid d \in D,\ d.a_d = a_I\}$. The support measures the applicability of the rule, and the confidence measures the association level with the outcome, which is the proportion of instances causing a desired outcome in the satisfying instances:

$$Conf(C \rightarrow O) = |D(C) \cap D(O)| / |D(C)|,$$

where $D(O) = \{d \mid d \in D,\ d.y = \text{true}\}$. Notice that the confidence only measures the outcome of the satisfaction cases, and the consequence of violating the constraint is not considered, which is also critical to evaluate the effectiveness of the constraint. Furthermore, the significance of the mined rules, which is very important for clinical evidence to be accepted, cannot be ensured just using support and confidence.

To address these problems, we use relative risk (RR) and Fisher's exact test to evaluate a constraint. Relative risk is the ratio of the probability of developing positive outcomes in an satisfaction group to the probability of developing positive outcomes in a comparison, violation group:

$$RR(C \rightarrow O) = \frac{Conf(C \rightarrow O)}{Conf((\neg C \cap C_I) \rightarrow O)}.$$

Relative risk has been widely used to quantify how strongly the presence or absence of a condition is associated with an outcome in clinical trials and risk analysis [14], which also has a clear interpretation in our scenario: "satisfaction cases of the constraint would be *RR* times as likely as violation cases to develop a positive outcome." A higher *RR* (> 1) means that the positive outcome is more likely to occur in the satisfaction group than in the violation group. Right-tailed Fisher's exact test can be performed to test if *RR* is significantly greater than 1. Fisher's exact test is valid for small and highly unbalanced dataset, which can therefore be applied to the rules with high support and few violation cases.

Another issue is that in real-world EMR, some of the data can be highly censored. The features with too many null values may cause an inaccurate calculation of support and relative risk. To tackle this problem, we replace the instance set $D(C_I)$ with $D(\hat{C}_I) = D(C_I) \cap \{d \mid d \in D, c \in C, d.x_c \neq \text{null}\}$ in the above formulas to obtain an unbiased evaluation of a rule. We also measure the valid rate of a rule, which is the proportion of instances without null value features:

$$VR(C) = |D(\hat{C}_I)| / |D(C_I)|.$$

Algorithm 1. Care Pathway Constraint Mining

Input: A draft care pathway (g, A, B) and a set of activities of interest A_I
The activity dataset D
Pre-defined thresholds: $\Theta = (\theta_V, \theta_S, \theta_R, \theta_P)$
Output: A set of constraints Φ

1: $\Phi = \emptyset$
2: **for each** $a_I \in A_I$ **do**
3: $L_1 = \emptyset$, $C_I = \{a_d = a_I\}$
4: **for each** $f \in A \cup B$ **do**
5: L_f = genCandidate1Rules(f)
6: **for each** $C_f \in L_f$ **do**
7: **if** evalRule(C_f, C_I, D, Θ) = true **then** $L_1 = L_1 \cup \{C_f\}$
8: **end for**
9: **end for**
10: **for** $k = 2$; $L_{k-1} = \emptyset$; $k = k + 1$ **do**
11: $L_k = \emptyset$, L_T = genCandidateRules(L_{k-1})
12: **for each** $C_T \in L_T$ **do**
13: **if** evalRule(C_T, C_I, D, Θ) = true **then** $L_k = L_k \cup \{C_T\}$
14: **end for**
15: **end for**
16: $\Phi = \Phi \cup$ {getBestRule($\cup_k L_k$, C_I, Θ)}
17: **end for**

As shown in Algorithm 1, our constraint mining algorithm takes as input the draft pathway model, the activity dataset derived from the EMR, as well as a set of pre-defined thresholds $\Theta = (\theta_V, \theta_S, \theta_R, \theta_P)$, where θ_V is the minimum valid rate θ_V, θ_S is the minimum support, θ_R is the minimum relative risk, and θ_P is the maximum p-value.

For each $a_I \in A_I$, the algorithm first builds a set of 1-rules L_1, where each rule has only one item (line 3-9). The function *genCandidate1Rules* (line 5) generates a set of candidate 1-rules for a feature *f*. If *f* is a categorical or Boolean feature, we create a set of candidate rules $L_f = \{C_f \mid C_f = \{(f, \omega_f, x_f)\},\ \omega_f \in \Omega_f,\ x_f \in X_f\}$, where X_f is the set of possible categories of *f*, and Ω_f is the possible operators (e.g., $\{KA = \text{false}\}$). If *f* is a numeric feature, we first discretize the attribute [12] and then build an item set similarly (e.g., $\{CR \leq 2.0\}$). Notice that $|X_f| \times |\Omega_f|$ candidate 1-rules are created for *f*, and not all of them are valid, frequently satisfied and correlated with the outcome. Therefore, we use the function *evalRule* to evaluate every rule $C_f \in L_f$ using the above metrics, and prune the uninteresting rules (line 7). For an event feature $f \in A$, if $VR(C_f \cup C_I) \geq \theta_V$ and $Sup(C_f \cup C_I) \geq \theta_S$, then C_f can be added to L_1. And for an condition feature $f \in B$, besides *VR* and *Sup*, we also check whether $RR(C_f \cup C_I \rightarrow O)$ is greater than θ_R to further remove the rules that are not very likely to cause positive outcomes.

In each subsequent pass *k*, the algorithm builds a set of *k*-rules L_k, where each rule has *k* items (line 10-15). The function *genCandidateRules* is similar to the Apriori algorithm [13], which uses L_{k-1} found in the pass *k*-1 to generate the candidate rules L_T by joining and pruning (line 11). Because one feature may have multiple candidate rule items with different operators and/or different reference values, the function also ensures that in each rule there is no implied items (e.g., $CR \leq 2$ and $CR \leq 3$) or conflicted items (e.g., $CR \leq 2$ and $CR > 3$). Then *evalRule* is performed to evaluate every rule $C_T \in L_T$ and remove the invalid, infrequent or negative rules (line 13).

After generating all the rules for a_I, the algorithm selects the best rule (line 16) that has as many items as possible and is significantly more likely to lead to a positive outcome. The function *getBestRule* first ranks the rules $C \in \cup_k L_k$ by *k* and *RR* orderly, and removes the rules whose p-values from Fisher's exact test is less than θ_P. Then the top-one rule in the rank list is selected as the constraint of a_I and added to the constraint set Φ. After all the iterations (line 2-17), the set Φ contains the discovered constraints for all the activities of interest in A_I.

Results and Discussion

We performed a case study using our method on real-world EMR collected from a China hospital. The EMR data contains 729,027 events (including 163,654 diagnoses, 178,019 medications and 387,354 lab tests) that occurred over 4 years, from a cohort of 2,693 patients with type 2 diabetes, where 1,454 patients also had cardiovascular diseases (CVD).

Two different care pathways were focused on in this study: (I) care pathway for the management of general type 2 diabetes; (II) care pathway for managing diabetes patients with CVD. The draft pathways are derived from a clinical guideline of type 2 diabetes mellitus [15] and recommendations from diabetes experts. In the care pathways, we defined clinical activities including 9 classes of medications and 7 panels of laboratory and vital sign tests, as well as attributes including 6 categories of problems and 12 test items. Some of the activities and attributes are shown in Table 2.

Table 2 – Abbreviations of activities and attributes

Abbr.	Clinical activity	Abbr.	Attribute
	Medications:		**Problems:**
BG	Biguanides	HG	hypoglycemia
SU	sulfonylureas	KA	ketoacidosis
αGI	α-glucosidase inhibitors		**Test Items:**
TCM	traditional Chinese meds	TG	triglyceride
ST	Statins	LDL	low density lipoprotein cholesterol
FI	Fibrates	HDL	high density lipoprotein cholesterol
	Laboratory tests:	ALT	alanine transaminase
MFT	metabolic function test	TB	total bilirubin
LFT	liver function test	CR	serum creatinine
RFT	renal function test	ACR	urine albumin/creatinine ratio

We first mined constraints for the general diabetes pathway (I). Its goal is controlling glucose levels (A1c < 7.0 %) while avoiding hypoglycemia (FPG < 3.9 mmol/L). The activities of interest are three classes of most widely used oral hypoglycemic drugs, as well as traditional Chinese antidiabetic medicines which are also used in China hospitals. The observation window and evaluation window were set to w_O = 1 year and w_E = 3 months, and the thresholds in Algorithm 1 were set to $\theta_V = 0.6$, $\theta_S = 0.75$, $\theta_R = 0.85$ and $\theta_P = 0.05$. The attributes used to evaluate outcomes were removed from the feature set to avoid the obvious correlation. Table 3 shows the number of instances for each activity, the mined constraint with temporal (on) and data (if) conditions, as well as its support and relative risk. All the constraints are statistically significant (p-value < 0.05). The CMMN representation of the first constraint in Table 3 is shown in Figure 3, which means that "use biguanides only after checking metabolic function, renal function and liver function, and only if hypoglycemia and ketoacidosis did not occur before, as well as CR ≤ 2.0 mg/dL, ACR ≤ 400 μg/mg and ALT ≤ 50 IU/L".

Table 3– Constraints mined for (I) (θ_V=0.6, θ_S=0.75, θ_R=0.85)

Act.	#Ins.	Constraint	Sup	RR
BG	4364	**on** *MFT*∧*RFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*CR*≤2∧*ACR*≤400∧*ALT*≤50	0.76	1.21
SU	2338	**on** *MFT* ∧*RFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤400∧*ALT*≤70∧*TB*≤30	0.92	1.39
αGI	2825	**on** *MFT*∧*RFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤500∧*ALT*≤50	0.93	1.57
TCM	201	**on** *MFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*ALT*≤50	0.94	1.63

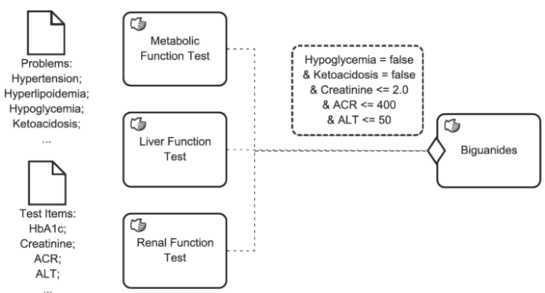


Figure 3 – Graphical representation of the constraint on BG

The discovered constraints are reasonable and meaningful, which indicate that all these medicines cannot be given to hypoglycemia or ketoacidosis patients, which is consistent with the clinical guideline [15]. Moreover, it is commonly accepted that severe hepatopathy and nephropathy are contraindications to the use of most oral antidiabetic drugs. Our study not only revealed that renal function and liver function should be checked prior to oral hypoglycemic medications, but also found the reference thresholds of the relevant test indicators for each pharmacy class, which are not always explicitly given in guidelines. For example, a very high CR and ACR usually denotes significant kidney diseases, and an obvious rise of ALT and TB normaly indicates severe liver diseases. The discovered constraints restrict that the medicines should not be used under these conditions.

We also extracted constraints for the care pathway (II) for diabetes with CVD. It has two stages: Stage 1 is similar with the pathway (I) for controlling glucose; and Stage 2 is designed to control lipid levels (TG < 1.7 mmol/L, LDL < 2.6 mmol/L and HDL > 1.0 mmol/L), which are critical indicators for CVD patients. The activities of interest in Stage 2 are two classes of medications to lower cholesterol: ST and FI. We used the dataset of the cohort of diabetes with CVD, and the parameters were the same with those in the above experiment. The discovered constraints are shown in Table 4. There is no constraint discovered for TCM and FI due to too few instances involved. Notice that the constraints in the pathway (II) are stricter than those in (I). That is because the disease condition of (II) is more severe and complex, and the interventions should be given more carefully.

Table 4– Constraints mined for (II) (θ_V=0.6,θ_S=0.75,θ_R=0.85)

Act.	#Ins.	Constraint (on *MFT*∧*RFT*∧*LFT*)	Sup	RR
BG	2326	**if** ¬*HG*∧¬*KA*∧*CR*≤2∧*ACR*≤70∧*ALT*≤60∧*TB*≤40	0.75	1.36
SU	1276	**if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤100∧*ALT*≤70∧*TB*≤30	0.91	1.41
αGI	1642	**if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤300∧*ALT*≤40∧*TB*≤30	0.83	1.42
ST	1624	**if** *CR*≤2∧*ACR*≤300∧*ALT*≤60∧*TB*≤30	0.75	1.56

The effect of varying the thresholds in Algorithm 1 was shown in Table 5. Compared with the constraints shown in Table 3, simpler constraints were extracted when using the higher thresholds, while more complex constraints can be discovered when using the lower thresholds. Therefore, clinicians can adjust the thresholds to obtain a desired constraint set, depending on how strict the care pathway is expected to be.

Table 5 – Constraints mined for (I) using different thresholds.

Thres.	Act.	Constraint	Sup	RR
θ_I=0.9	*BG*	**on** *MFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*ALT*≤50	0.93	1.40
θ_S=0.9	*SU*	**on** *MFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*ALT*≤70	0.97	1.89
θ_R=1.0	*αGI*	**on** *MFT*∧*LFT* **if** ¬*HG*∧¬*KA*∧*ALT*≤50	0.96	1.29
θ_I=0.5 θ_S=0.5 θ_R=0.7	*BG*	**on** *MFT*∧*RFT*∧*LFT* (the same below) **if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤400∧*ALT*≤60∧*TB*≤40	0.91	1.34
	SU	**if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤400∧*ALT*≤70∧*TB*≤30	0.92	1.39
	αGI	**if** ¬*HG*∧¬*KA*∧*CR*≤3∧*ACR*≤90∧*ALT*≤40∧*TB*≤30	0.78	1.30

We also compared our algorithm with the existing association rule mining algorithm on mining constraints for the pathway (I). For each activity, we applied the CBA-RG algorithm [12] to extract the rules, and selected the rule with the most items and the highest *Conf* as the constraint. We tried different thresholds for the CBA-RG algorithm, and show the results of *minsup* = 0.75 and *minconf* = 0.7 in Table 6, which are the most reasonable on the whole. Notice that some discovered conditions are less meaningful (e.g., CR ≤ 4 is too loose for restricting BG, while ACR ≤ 30 is over strict for using *αGI*), which is because the consequence of violating a constraint are not considered in the CBA-RG algorithm.

Table 6 – Constraints mined for (I) using Sup and Conf

Act.	Constraint (on *MFT*∧*LFT*∧*RFT*)	Sup	Conf
BG	**if** ¬*HG*∧¬*KA*∧*CR*≤4∧*ACR*≤60∧*ALT*≤40∧*TB*≤40	0.78	0.70
SU	**if** ¬*HG*∧¬*KA*∧*CR*≤2∧*ACR*≤300∧*ALT*≤60∧*TB*≤30	0.77	0.71
αGI	**if** ¬*HG*∧¬*KA*∧*CR*≤4∧*ACR*≤30∧*ALT*≤60∧*TB*≤90	0.76	0.74

The proposed approach was implemented using Java, and the execution of each experiment above was completed in several minutes on a laptop computer. The approach was integrated into IBM Care Pathway Workbench [16], which is a web-based platform to build care pathways. Each discovered constraint can be translated into an entry criterion in a CMMN-based care pathway model [9], and clinicians can further edit and refine the constraints in the workbench. Figure 4 shows a screenshot of the workbench, demonstrating the structured representation of the above constraint shown in Figure 3. The final care pathway models can then be loaded to a care management system to provide evidence-based decision support during individual patient care.

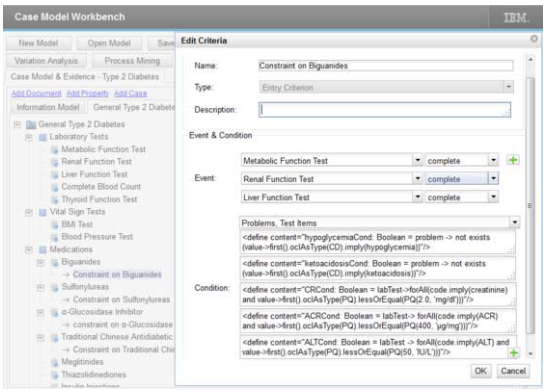


Figure 4 – The constraint in Care Pathway Workbench

Conclusion

Care pathway mining [4,5], which is aimed at discovering the frequent and best practices that impact patients in their care journeys, can provide meaningful insights into care pathway modeling and execution. Though the temporal and data preconditions of clinical activities forms the most critical part of a care pathway, it is still a challenging problem to mine the constraints that can represent the best practices for a specific group of patients. In this paper, we proposed a care pathway mining method to extract temporal and data constraints that are correlated with positive outcomes from EMR. We performed a real-world case study using our method on different cohorts of diabetes patients, and the results showed that our method can discover reasonable constraints for care pathways, which has the potential to provide meaningful decision support during care pathway execution.

One limitation of our method is the scalability of the constraint mining algorithm. Though the experiments showed that the algorithm worked well for care pathways with dozens of features and for relatively high thresholds, if it was applied to much more features and very low thresholds, its running time and space requirement might dramatically increase. In future work, a scalable algorihm with parallel computing could be developed to improve the runtime performance. Another limitation is that the activities, attributes, possible operators, and thresholds have to be manually defined before constraint mining. Potential improvements for this problem include to automatically identify activities and attributes of interest from candidate features in EMR by feature selection, and to learn possible operators and thresholds from user-defined sample constraints. Besides, our current method cannot extract the exit criteria of activities (e.g., stop using fibrates after initiating statins), which are also common and important in pathways. In future work, we could modify the mining algorithm to support the discovery of the constraints for terminating activities that are prone to cause negative outcomes in specific conditions.

References

[1] Vanhaecht K, De Witte K, Sermeus W. The impact of clinical pathways on the organisation of care processes. Belgium: Katholieke Universiteit Leuven; 2007.

[2] Rebuge A, Ferreira DR. Business process analysis in healthcare environments: a methodology based on process mining. Information Systems. 2012; 37(2): 99-116.

[3] Lang M, Bürkle T, Laumann S, Prokosch HU. Process mining for clinical workflows: challenges and current limitations. Stud Health Technol Inform. 2008;136:229-34.

[4] Huang Z, Lu X, Duan H and Fan W. Summarizing clinical pathways from event logs. J Biomed Inform. 2013; 46: 111-27

[5] Lakshmanan GT, Rozsnyai S, Wang F. Investigating clinical care pathways correlated with outcomes. Business Process Management. 2013: 323-38.

[6] Bose RPJC, Maggi FM, van der Aalst WMP. Enhancing declare maps based on event correlations. Business Process Management. 2013: 97-112.

[7] Maggi FM, Dumas M, García-Bañuelos L,~ ~ ~ ~ ~~ . Discovering data-aware declarative process models from event logs. Business Process Management. 2013: 81-96.

[8] Patkar V, Fox J. Clinical guidelines and care pathways: a case study applying PROforma decision support technology to the breast cancer care pathway. Stud Health Technol Inform. 2008; 139: 233-42.

[9] Mei J, Li J, Yu Y, Li X, Liu H, Xie G. Embracing case management for computerization of care pathways. Stud Health Technol Inform. 2014; 205: 3-7.

[10]Li X, Liu H, Zhang S, Mei J, Xie G, Yu Y, et al. Automatic variance analysis of multistage care pathways. Stud Health Technol Inform. 2014; 205: 715-9.

[11]Object Management Group. Case management model and notation (CMMN). [released 2014 May; cited 2014 Dec 17]. Available from: http://www.omg.org/spec/CMMN/.

[12]Liu B, Hsu W, Ma Y. Integrating classification and association rule mining. Proceedings of KDD. 1998: 80-6.

[13]Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of VLDB. 1994: 487-99.

[14]Li J, Fu AW, Fahey P. Efficient discovery of risk patterns in medical data. Artif Intell Med. 2009; 45(1): 77-89.

[15]Chinese Diabetes Society. China guideline for type 2 diabetes. Peking University Medical Press; 2011.

[16]Yu Y, Liu H, Li J, Li X, Mei J, Xie G, et al. Care pathway workbench: evidence harmonization from guideline and data. Stud Health Technol Inform. 2014; 205: 23-7.

Address for correspondence

Xiang Li, PhD. IBM Research – China, Building 19, Zhongguancun Software Park, Haidian District, Beijing, China, 100193.

Email: lixiang@cn.ibm.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-716

Acquiring Plausible Predications from MEDLINE by Clustering MeSH Annotations

Jose Antonio Miñarro-Giménez, Markus Kreuzthaler, Johannes Bernhardt-Melischnig, Catalina Martínez-Costa, and Stefan Schulz

Institute of Medical Informatics, Statistics, and Documentation, Medical University of Graz, Austria

Abstract

The massive accumulation of biomedical knowledge is reflected by the growth of the literature database MEDLINE with over 23 million bibliographic records. All records are manually indexed by MeSH descriptors, many of them refined by MeSH subheadings. We use subheading information to cluster types of MeSH descriptor co-occurrences in MEDLINE by processing co-occurrence information provided by the UMLS. The goal is to infer plausible predicates to each resulting cluster. In an initial experiment this was done by grouping disease-pharmacologic substance co-occurrences into six clusters. Then, a domain expert manually performed the assignment of meaningful predicates to the clusters. The mean accuracy of the best ten generated biomedical facts of each cluster was 85%. This result supports the evidence of the potential of MeSH subheadings for extracting plausible medical predications from MEDLINE.

Keywords:

Knowledge Acquisition; MEDLINE; Biomedical Terminologies.

Introduction

The availability of massive datasets in the biomedical domain presents challenges and opportunities for data management and analysis. One of the most established sources is the literature database MEDLINE[1] with over 23 million bibliographic records. One of its most valuable assets is metadata annotations performed by domain experts at the U.S. National Library of Medicine, using the MeSH[2] indexing vocabulary.

Such annotations consist of unordered lists of so-called headings, i.e., controlled terms from the MeSH vocabulary, most of which are further qualified by subheadings (Table 1).

Table 1- MeSH annotations of a sample paper (ID= 1341068)

MeSH Heading with corresponding concept IDs (UMLS CUI)	MeSH Subheading with abbreviations
Gastritis (C0017152)	Etiology (ET)
Gastritis (C0017152)	Physiopathology (PP)
Gastrointestinal Diseases (C0017178)	Etiology (ET)
Gastrointestinal Diseases (C0017178)	Physiopathology (PP)
Helicobacter Infections (C0079487)	Etiology (ET)
Helicobacter pylori (C0079488)	Physiology (PH)

The subheading information is not relational; subheadings only qualify the subject of what could be a complete statement such as, "Helicobacter pylori causes gastritis."

These annotations are available via the UMLS [1] Metathesaurus (UMLS Meta), a resource that connects over 9 million terms for some 2.8 million concepts from more than 130 biomedical terminology systems[3]. It provides a file named MRCOC5[4], which contains MeSH annotation co-occurrences. This data source has repeatedly been used to extract concept associations [2-4], such as between genes and diseases [5], as a support for text mining [6], for word sense disambiguation [7], and query expansion [8].

The large amount of data to be processed requires efficient and scalable methods related to big data analysis [9]. Such methods entail a distributed framework of multithread, massively parallelized computing tasks.

Table 2- Possible predicates between concepts ordered by selected semantic types (columns: subjects; rows: objects)

	Disease/ Syndrome	**Finding**	**Substance**	**Organism**
Disease/ Syndrome	complicates causes co-occurs with	occurs in diagnoses	treats prevents causes occurs in	affected by causes
Finding	produces diagnosed by	complicates causes co-occurs with	treats prevents causes	affects caused by
Substance	caused by treated by prevented by diagnosed by	treated by caused by prevented by	interacts	is affected by produces
Organism	caused by affected by	caused by affects	affects produced by	interacts with

Here we describe the first experiments to investigate the potential of MeSH annotation co-occurrence data from MEDLINE, in concert with machine learning methods and additional terminological resources. This is a novel approach compared to existing methodologies that use similar resources. The main goal is to extract concept-concept predications based on statistical associations regarding the distribution of 83 subheading types used in MeSH annotations. We understand by predications subject-predicate-object triplets, which convey

[1] http://www.nlm.nih.gov/pubs/factsheets/medline.html
[2] http://www.nlm.nih.gov/mesh/
[3] http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html
[4] http://mbr.nlm.nih.gov/MRCOC/MRCOC_Doc_2013.pdf

information that is not ontological in the strict sense [10,11] of formal, logic-based ontology, but stands for a strong or typical association between the subject and the object concept in terms of the predicate. For example, the triplet

<gastritis; has-location; gastric mucosa>

is ontological, as it can be interpreted that all instances of gastritis are located in some gastric mucosa, whereas

<gastritis; caused-by; helicobacter pylori>

cannot be translated in logics. The interpretation is less clear; it might be paraphrased as: "It is sufficiently frequent that gastritis is caused by helicobacter pylori for this to be considered in a clinical decision-making process".

We focus our experiments on the MeSH headings associated with the semantic types *Disease/Syndrome* and *Pharmacologic Substance* as defined in UMLS Semantic Network (UMLS SN). Table 2 lists possible predicates between semantic types as defined in UMLS SN. They represent probable predicates that can relate concepts from UMLS Meta that belong to such semantic types. Furthermore, we limit our dataset to MEDLINE records published in the past five years to simplify the experiments.

Materials and Methods

MRCOC provides pairs of concepts that co-occur in the same entry of an information source. It summarizes the MeSH descriptors that occur together in MEDLINE citations from the MEDLINE/PubMed baseline, a snapshot created at the beginning of each new MeSH indexing year. In this project we have used the version of MRCOC from the UMLS Meta 2014 AA release. Besides the main headings, MRCOC includes a "fingerprint" of MeSH subheadings, such as "DT" (Drug Therapy) or "PC" (Prevention & Control). They characterize the context in which the first concept occurs in the related MEDLINE records. For example, if an article is about the prevention of stroke, the concept "Stroke" in the MEDLINE record is refined by the subheading "PC".

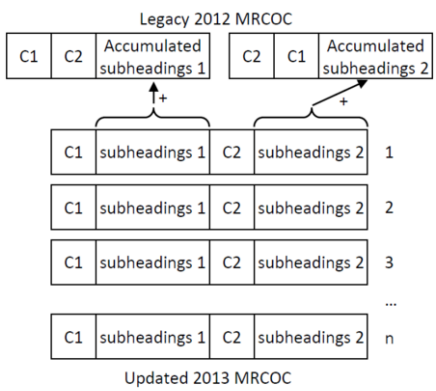


Figure 1 - It shows the content of the legacy and updated versions of the MRCOC file regarding the co-occurrences of C1 and C2 concepts and their corresponding MeSH subheadings.

Since 2013, an updated version of the MRCOC file (size 131GB)[5] is provided. Figure 1 shows how the content of the legacy and the updated version of MRCOC have changed. In the updated version, each line contains a pair of two concepts that co-occur in MEDLINE with their corresponding MeSH subheadings concepts. On the other hand, the legacy version had accumulated the MeSH subheadings of every co-occurrence in two lines, one with the pair of co-occurring concepts and the accumulated subheadings related to the first concept, and another with the co-occurring concepts but with the accumulated subheadings of the second concept. Consequently, the accumulation of MeSH subheadings now has to be computed out of the updated MRCOC file. In MRCOC, the MeSH descriptors are represented as UMLS concepts, identified by concept identifiers (CUIs), and the subheading information as two-letter abbreviations (see Table 1). The UMLS semantic types and relations (used as a source of candidate predicates like in Table 2) were obtained from the files MRSTY and SRSTRE1.

The methodology to extract plausible predications from MEDLINE records consists of the following sequential steps: (1) aggregate the co-occurring concept pairs and their subheading information; (2) obtain the corresponding log-likelihood rates (LLRs); (3) cluster the co-occurring concept pairs based on the vector of accumulated MeSH subheading values; and (4) manually associate meaningful predicates to the resulting clusters.

Aggregate co-occurring concept pairs

Due to the huge volume of data in our project we applied the MapReduce [12] programming paradigm and the Amazon cloud services (Amazon Elastic MapReduce[6], Amazon S3[7]) together with Apache Hadoop [13] for generating the aggregated version of MRCOC, as depicted by Figure 2. MapReduce has two procedures: *Map*(), which filters and sorts the elements to be processed; and *Reduce*(), which takes the resulting elements from *Map*() and performs operations on the accumulated values to the unique key. Input and output data are represented as of key/value pairs, thus facilitating data access and distribution across several computers. Hadoop is a software framework that supports the distributed processing of large data sets across clusters of computers. As a result, jobs can be easily scaled up and parallelized. We therefore implemented four JobFlowSteps within a JobFlow executed on Amazon Elastic MapReduce. The first job filtered (past five years, ZY-Yes) the detailed_CoCoccurs_2014.txt.gz file (split and uploaded to Amazon S3) and obtained the aggregated values for co-occurring concept pairs.

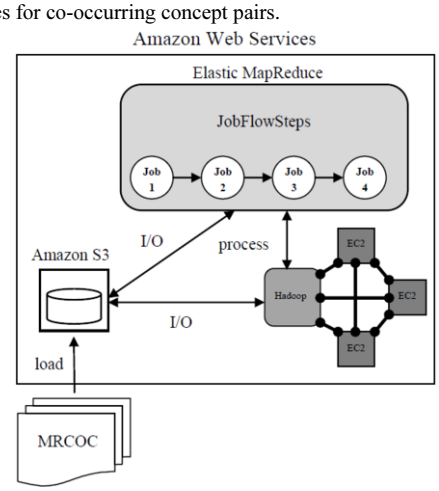


Figure 2 - System architecture in Amazon cloud services of the processing of MRCOC file and the calculation of the corresponding likelihood.

[5] http://mbr.nlm.nih.gov/MRCOC.shtml

[6] http://aws.amazon.com/elasticmapreduce/?nc2=h_ls

[7] http://aws.amazon.com/s3/?nc2=h_ls

Obtaining LLRs

The log-likelihood rates (LLRs) are calculated to filter co-occurring concepts that are not significant or without sufficient scientific impact. The LLRs were determined as part of the MapReduce JobFlow exploiting the Apache Mahout[8] implementation along with the aggregation of MRCOC. Thus, the *Map()* procedure populates the intermediate key/value pairs by using the first concept (CUI1) combined with the second concept (CUI2) as the key, the MeSH subheading abbreviations from CUI1 as the value, and, inversely, CUI2 combined with CUI1 as the key and the MeSH subheading abbreviations from CUI2 as the value. The reduce phase generates the accumulated MeSH subheading vector per key "CUI1CUI2" along with the number of times CUI1 and CUI2 co-occur (#CUI1CUI2). The second job takes the output from the first job and computes the absolute number of occurrences for CUI1 (#CUI1) and CUI2 (#CUI2).

The third job performs a reduce-side join on CUI1, adding the absolute number of occurrences for CUI1 to the output from the first step, with respect to "CUI1CUI2". The last job once again applies a reduce-side join but this time on CUI2, adding the number of absolute occurrences for CUI2 to the co-occurrence under inspection. Now in this final JobFlowStep missing parts for calculating the LLR (see Table 3) are resolved (#notCUI1notCUI2 #CUI1notCUI2 #notCUI1CUI2), and finally the LLR is added to the accumulated subheading information.

Table 3. It shows a summary of the relations between the different counts calculated from the concept co-occurrences.

Co-occurrences	CUI1	not occur CUI1
CUI2	#CUI1CUI2	#notCUI1CUI2
not occur CUI2	#CUI1notCUI2	#notCUI1notCUI2

Formula (1) shows how the log-likelihood ratio (LLR) [14] is computed:

$$LLR = 2 \times \left(H(matrix) - H(mRows) - H(mCols) \right) \quad (1)$$

The function H is the Shannon entropy [15]. *H(matrix)* corresponds to the entropy of #CUI1CUI2, #notCUI1notCUI2, #CUI1notCUI2, and #notCUI1CUI2. *H(mRows)* is the sum of entropies of the pairs (#CUI1CUI2, #notCUI1CUI2) and (#CUI1notCUI2, #notCUI1notCUI2). *H(mCols)* is the sum of entropies of the pairs (#CUI1CUI2, #CUI1notCUI2) and (#notCUI1CUI2, #notCUI1notCUI2). Applying a threshold of 10.83 (chi-squared test, f=1, p<0.001) to the calculated LLR in the last job, we obtain the list of relevant co-occurrences that are used for the clustering approach.

Co-occurrence clustering

In this phase, the resulting list of co-occurrences is grouped based on the accumulated list of MeSH subheadings. The clustering was performed using WEKA [16], which provides a set of machine learning algorithms mainly for data mining tasks. We chose the simple k-means clustering algorithm. In order to decide the types of predicates to assign to the co-occurring concept pairs we first identify the UMLS semantic types of each concept. In this experiment, we have focused on the predications that relate concepts from the semantic types *Disease/Syndrome* (T047) and *Pharmacologic Substance* (T121). According to UMLS SN, concepts under these semantic types can be related via the predicates (SN relations) *diagnoses, affects, treats, complicates, prevents,* and *causes* (Figure 3).

[8] http://mahout.apache.org/

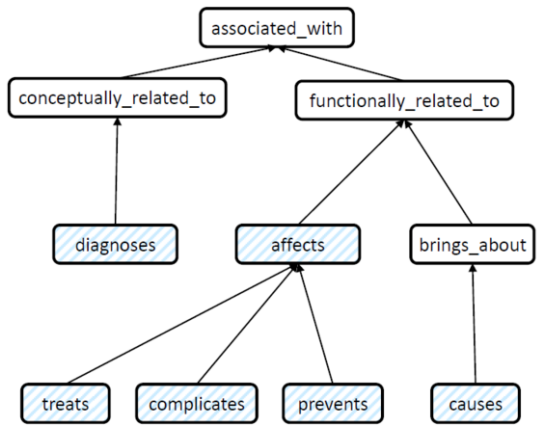


Figure 3 - Excerpt of the UMLS Semantic Network related to the predicates (UMLS SN relations): diagnoses, affects, treats, complicates, prevents, and causes.

We used the number of suggested predicates as the parameter of the number of clusters to be produced by the k-means algorithm. The preparation of the co-occurring concept pairs involves the division of the list into two, the first list containing the co-occurrences *Disease/Syndrome* concepts with *Pharmacologic Substance* concepts, whilst the second list contains the co-occurrences in the inverse order. The reason is that in the original record, subheadings are only assigned to the first (subject) concept. For example, in C_1;C_2;DT, the subheading DT ("drug therapy") refers to C_1, in C_2;C_1;TU, the subheading "therapeutic use" (TU) belongs to C_2. Besides, to normalize the frequency of MeSH subheadings for each co-occurrence we divided them by the total times the pair of concepts co-occurs. Consequently, the normalized values are in the range 0-1.

Cluster association

This final phase connects resulting clusters to some of the suggested predicates. The matching of predicates to cluster and the evaluation of the plausibility of the resulting predications is a manual task to be performed by domain experts. To facilitate the association, we selected the ten best co-occurrences of each cluster, i.e., the ten nearest co-occurrences to each cluster centroid based on their Euclidean distance from the normalized MeSH subheading vector to the centroids subheading vector. The experts evaluated the plausibility of each collected co-occurrence with the matched predicate of the cluster to estimate the accuracy of the inference. The distance of the co-occurrences to their corresponding cluster centroid and the accuracy of the associated predicate in the cluster indicate the consistency of the inferred biomedical fact.

Results

The main result of the project is the definition and implementation of the methodology to extract meaningful Subject-Predicate-Object triplets through the processing and clustering of statistically associated concepts from MEDLINE records.

The method was applied only to the predications between diseases and substances, limited also to the years 2009 – 2013, where the two concepts are flagged major topics (Flag ZY). These co-occurrences were processed and filtered using the system described in Figure 2. As a result, we obtained a list of 17,908 co-occurrences such as "Addison's Disease" and "Cortisol", or "Osteoporosis" and "Calcium".

The UMLS Semantic Network identifies six different predicates (SN relations) between the semantic types *Disease/Syndrome* and *Pharmacologic Substance*. Therefore, the clustering process grouped the list of co-occurrences into six different clusters.

The clustering is based on the subheadings vector of each concept pair. Consequently, some subheadings are more predominant in one cluster and other subheadings in other clusters. Table 4 shows the number of different co-occurrences in each cluster.

Table 4 - Total number of co-occurrences in each resulting cluster.

Cluster	Disease/Syndrome vs. Pharmacol. Substance
0: Diagnosis and Etiology	2,422
1: Chemically Induced	2,015
2: Prevention and Control	1,741
3: Drug Therapy	7,668
4: Complications and Therapy	2,543
5: Metabolism and Pathology	1,519

The clustering process creates the clusters that minimize the distance of co-occurrences based on their MeSH subheading vectors and indicates the nearest cluster centroid to each co-occurrence. This means that the co-occurrences are associated to only one cluster. This is a limited approach since many drugs can both be used to prevent and to treat a disease. However, we can calculate the distance between a co-occurrence and each other cluster and obtain the plausibility of also being close to a second or third cluster. Moreover, we can calculate the mean values of two cluster centroids and check whether the joint cluster represents co-occurrences related to both predicates at the same time.

The manual association of a predicate to each cluster by a domain expert produced a list of plausible SPO triplets. To evaluate these triplets, we collected the elements nearest to each cluster centroids and manually analyzed their accuracy. Because this manual evaluation is a time consuming task we limited it to the ten best elements of each cluster. Consequently, the domain experts only evaluated 60 plausible triplets. The manual association and evaluation of the predicates with clusters provided the following results:

- Cluster 0 with the predicate *diagnoses* gave an accuracy of 50% and the mean of the distance values for the ten nearest co-occurrences is 0.3537.
- Cluster 1 and the predicate *causes* were manually associated. The resulting accuracy of the association is 90% and the mean of the distance values is 0.2136 for the top ten co-occurrences in the cluster.
- Cluster 2 was linked to the *prevents* predicate. It has an accuracy of 100% for the ten nearest co-occurrences in cluster 2. The mean value of the distances of the top ten co-occurrences is 0.2590.
- Cluster 3 with the *treats* predicate also has an accuracy of 100%. The ten nearest co-occurrences were manually examined and associated to the *treats* predicate. The mean of the distance values of the best co-occurrences in the cluster is 0.1621.
- Cluster 4 was related to the *affects* predicate. The inspected co-occurrences were not clearly related to a more specialized relationship and *affects* was general enough to cover them. Therefore, we obtained an accuracy of 100% but a mean distance value of 0.3194.
- Cluster 5 was linked to the *causes* predicate. The calculated accuracy is 70% and the mean of the distance values of the top ten co-occurrences is 0.2440.

Discussion

The association of predicates to co-occurrences presents some difficulties. Many of the concepts that co-occur represent general terms and, therefore, the acquired facts represent coarse-grained knowledge. For example, the co-occurrence in cluster 0 “Drug Allergy” and “Anticoagulants” are related to the predicate *causes*, but in fact, any drug has the potential to cause drug allergy. Furthermore, MeSH’s granularity is limited both regarding disorders and substances. Another problem is the ambiguity of co-occurrences from which more than one predicate can be inferred. For example, the co-occurrence in cluster 0 of the disease concept *Coagulation Defect* and the pharmacologic substance concept *Blood Coagulation Factor* could be linked either to *treats* if there is a substitution of the coagulation factor or to *diagnose*, because the lack of certain coagulation factor proteins in blood is used to diagnose corresponding coagulation defects. This ambiguity is reflected in the fact that the clusters have a higher distance value of their members to their centroids. We also found that certain associations do not correspond to any of the assigned predicates. For example, cluster 4 contains the co-occurrence of *Asthma* with *Insulin*, from the experimental treatment of patients with asthma using inhaled insulin [17]. This has not yet been universally accepted as a therapy, and its translation into a generalizable predicate would be problematic.

An important observation was that the most predominant MeSH annotations mostly agreed with the assigned predicate and, hence, they could be used to support the manual curation done by the physicians. For example, in cluster 3 the most predominant subheading was *Drug Therapy*, which corresponded quite well to the assigned predicate *treats*.

In general, we have observed a good correlation between the accuracy of the predicates and the distance values of co-occurrences, i.e., the lower the accuracy the higher the distance values to the centroid of the cluster. For example, the *treats* predicate, which exhibited an accuracy of 100% based on the nearest ten members of cluster 3, is related to more specialized concepts such as *Cerebrovascular accident* and *Alteplase*, and, thus, the distance to the centroid of cluster 3 is lower than in the other clusters. On the other hand, the cluster 0 has an accuracy value of 50% for the predicate *diagnoses* and the mean value of the distance of the top ten co-occurrences is 0.3537.

Conclusions and Further Work

MapReduce and cloud services have demonstrated the potential of big data analysis, with the processing of more than 11 million co-occurences in minutes, a process that would take hours in a centralized single-thread architecture. It will therefore be no problem to extend the analysis to the complete dataset, including all annotations, even inferred ones, and possibly extending it to additional features extracted from titles and abstracts.

Our methodology demonstrated its ability to extract biomedical facts from MEDLINE records. However, there was a focus on very general statements due to the predominance of coarse-grained MeSH headings. Consequently, the acquired knowledge may be trivial for experts but could be of interest to lay persons in medical domain. We plan to improve the specificity of the generated facts by including additional knowledge from

bibliographic records using Natural Language Processing methods that recognize terms in abstracts and map them to SNOMED CT and other more fine-grained terminologies. In this case, subheading information can still be used if one of the two concepts stems from the MEDLINE metadata.

More evaluation would be necessary to rate the plausibility or correctness of the extracted facts, in order to assess the suitability of our methods for different use cases and to detect weaknesses that could trigger methodological modifications or enhancements.

Considering that every pair of co-occurring concepts provides two different subheading vectors, because the subheadings associated with either concept are different, a more subtle distinction between the extracted facts would be possible. Other information could be obtained from the analysis of temporal trends, publication types, or the extraction of negations and other epistemic information (e.g., "does not influence", "no evidence of") from titles and abstracts.

References

[1] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;7:D267–D270.

[2] Mendonça EA, Cimino JJ. Automated knowledge extraction from MEDLINE citations. *Proc AMIA Symp*. 2000: 575-579.

[3] Bodenreider O, Aubry M, Burgun A. Non-lexical approaches to identifying associative relations in the gene ontology. *Pac Symp Biocomput*. 2005:91-102.

[4] Schulz S, Costa CM, Kreuzthaler M, Miñarro-Giménez J A, Andersen U, Jensen AB, Maegaard B. Semantic Relation Discovery by Using Co-occurrence Information. LREC 2014.

[5] Cantor MN, Sarkar IN, Bodenreider O, Lussier YA. Genestrace: phenomic knowledge discovery via structured terminology. *Pac Symp Biocomput*. 2005:103-114.

[6] Srinivasan P, Hristovski D. Distilling conceptual connections from MeSH co-occurrences. *Stud Health Technol Inform.* 2004;107(Pt 2):808-812.

[7] Agirre E, Soroa A, Stevenson M. Graph-based Word Sense Disambiguation of biomedical documents. *Bioinformatics*. 2010; 26(22):2889-2896.

[8] Otegi A, Arregi X, Agirre E. Query Expansion for IR using Knowledge-Based Relatedness. *In IJCNLP*. 2011: 1467-1471.

[9] Mohammed EA, Far BH, Naugler C. Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. *BioData Min*. 2014;7:22.

[10] Rector A. Barriers, approaches and research priorities for integrating biomedical ontologies. *SemanticHealth Deliverable* 6.1. Available from: www.semantichealth.org/DELIVERABLES/SemanticHEALTH_D6_1.pdf.

[11] Schulz S, Jansen L. Formal ontologies in biomedical knowledge representation. *Yearb Med Inform.* 2013;8(1):132-46.

[12] Dean J, and Ghemawat S. "MapReduce: simplified data processing on large clusters." Communications of the *ACM* 2008: 107-113.

[13] White T. Hadoop: the definitive guide: the definitive guide. *O'Reilly Media, Inc*., 2009.

[14] Dunning T. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics* 1993: 61-74.

[15] Shannon CE.A mathematical theory of communication. *Bell System Technical Journal*. 1948; 27: 379-423.

[16] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl*. 2009; 11 (1): 11-18.

[17] Mudaliar S, Henry RR. Inhaled insulin in patients with asthma and chronic obstructive pulmonary disease. *Diabetes Technol Ther.* 2007;9 (1):S83-92.

Acknowledgements

This work was performed as a part of the BMFacts project (BMFacts: Knowledge acquisition for a biomedical *fac*t repository), funded by the Austrian Science Fund (FWF): [M 1729-N15].

Address for correspondence

jose.minarro-gimenez@medunigraz.at

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-721

Effects of Plasma Transfusion on Perioperative Bleeding Complications: A Machine Learning Approach

Che Ngufor, Dennis Murphree, Sudhindra Upadhyaya, Nageswar Madde, Daryl Kor, and Jyotishman Pathak

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

Perioperative bleeding (PB) is associated with increased patient morbidity and mortality, and results in substantial health care resource utilization. To assess bleeding risk, a routine practice in most centers is to use indicators such as elevated values of the International Normalized Ratio (INR). For patients with elevated INR, the routine therapy option is plasma transfusion. However, the predictive accuracy of INR and the value of plasma transfusion still remains unclear. Accurate methods are therefore needed to identify early the patients with increased risk of bleeding. The goal of this work is to apply advanced machine learning methods to study the relationship between preoperative plasma transfusion (PPT) and PB in patients with elevated INR undergoing noncardiac surgery. The problem is cast under the framework of causal inference where robust meaningful measures to quantify the effect of PPT on PB are estimated. Results show that both machine learning and standard statistical methods generally agree that PPT negatively impacts PB and other important patient outcomes. However, machine learning methods show significant results, and machine learning boosting methods are found to make less errors in predicting PB.

Keywords:

Blood Transfusion; Bleeding; Treatment Effect; Classification; Machine Learning; Electronic Health Records.

Introduction

Bleeding in patients undergoing surgical procedures is a serious and relatively common complication that has been found to be associated with increased health care resource utilization, morbidity, and mortality. Although the origin of bleeding during surgery may be due to multiple factors, surgical factors and pre-existing abnormalities of the hemostatic system represent the principal causes of significant perioperative hemorrhage.[1] Excessive bleeding, reoperation for bleeding, and the need for transfusion of blood products are common both during and shortly after some types of surgical procedures. In spinal surgery, for example, between 30% and 60% of patients require allogeneic blood transfusion.[2] Reoperation for bleeding and administration of blood products are associated with postoperative complications including transfusion-associated lung injury (TRALI).[3] Despite the known negative effect on immunomodulation and increased risk of postoperative complications and mortality, various studies have shown large variability in the use of blood products among different centers, and even among individual anesthesiologists within the same center.[4, 5] For example, the number of red blood cell (RBC) units transfused annually in the US alone is about 14 million. Cost-wise, at an estimated cost of $761 per unit of RBC, this amounts to $10.5 billion in health care expenditures.[4] Therefore, strategies to minimize the need for allogeneic blood transfusions and perioperative bleeding complications are of great interest.

A key step to reduce the need for transfusion of allogeneic blood product is to assess patients who might bleed pre-transfusion. This might involve the identification of patients with impaired hemostasis and increase bleeding risk. However, our ability to predict these adverse events, typically estimated with a combination of patient- and procedure-related factors,[6] is incomplete. For patient-related factors, substantial emphasis is often placed on preoperative screening tests, such as the international normalized ratio (INR): a major driver for decisions about preoperative plasma transfusion.[7]

Fresh frozen plasma infusions are commonly used to improve coagulation or clotting and are the main therapy option for patients with elevated INR. A large proportion of the plasma components are transfused in the perioperative environment,[7,8] however, they are frequently administered prophylactically in the absence of significant active bleeding. This practice persists despite a growing body of literature questioning its efficacy.[8, 9] Moreover, plasma transfusions are increasingly recognized as important contributors to transfusion-related complications, including allergic reactions, TRALI, and transfusion-associated circulatory overload (TACO).[3, 10]

The goal of this study is to apply advanced machine learning methods to study the effect of preoperative plasma transfusion (PPT) on perioperative beeding (PB). The study is built on recent work described in Jia et al.[11] Based on propensity score matching estimated via standard statistical methods, Jia et al showed that PPT does not improve PB complications for patients with high INR scores. Generalized linear models such as logistic regression or linear regression are frequently used to estimate the propensity scores and models for the outcome. However, parametric models require assumptions regarding the functional form, distribution of the variables, and variable selection. If any of these assumptions are incorrect, the derived causal relationship may be misleading. Contrary to statistical approaches, machine learning methods can estimate complex relationships between the outcome and observed variables producing consistent estimates of the propensity scores and hence more reliable causal relationships.

Unlike the standard statistical analyses in Jia et al,[11] this work investigates the causal relationship between PPT and PB

by making use of modern machine learning algorithms to obtain consistent and reliable relationships. The relationship between PPT and other secondary outcomes such as intraoperative RBC transfusion (Intra-RBC), blood loss, Re-operation for bleeding (Re-OP), need for ICU care, ICU length of stay (ICU-LOS), hospital length of stay (LOS), and mortality are also studied. Different from the approach in Jia et al,[11] the analyses in this study are geared towards designing advanced techniques to identify early patients who might bleed during or after a surgical operation. In particular, in the interesting case where potential non-bleeding patients are identified, a more conservative practice can be adopted in which PPT is administered not only based on elevated INR, but on accurate predicted individual need for PPT. The presented work is an initial step towards achieving this larger goal. In this light, the problem is studied under the framework of causal inference by estimating the causal effect of the treatment (in this case PPT). Thus, estimates of the average treatment effect for those who received PPT can be used to quantify the independent association of PPT with PB. The computed average treatment effect can be regarded as a form of attributable risk, comparing the overall risk of bleeding to the risk of bleeding for those who received PPT.[12] Large values can be interpreted as increased risk of bleeding.

Although machine learning methods have been applied extensively in healthcare, to the best knowledge of the authors, their use for estimating the causal treatment effect of blood product transfusion has not been investigated. Thus, besides being one of the first works on the application of machine learning for blood transfusion, this paper strives to study PPT on several important patient outcomes in the context of causal inference by computing robust estimates of the average effect of PPT that can be used to answer questions like "what and how big is the effect of PPT on patient outcomes?"

Data Structure and Parameter of Interest

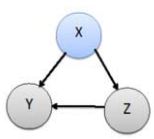


Figure 1: Causal Graph and Confounding

This study is based on an observational comparative effectiveness research analysis using a cohort design described in Jia et al [11] that was approved before initialization and this study followed all guidelines for strengthening and reporting of observational studies.

The structure of the data used in this study has the typical structure of causal inference. The observations for each patient are given by (**X, Y, Z**) where $Z \in \{0,1\}$ is the treatment indicator with **Z**=1 if patient was treated or **Z**=0 if patient was not treated. **X** is a vector of baseline covariates that records information specific to each patient prior to treatment. Y is the outcome variable such as perioperative bleeding, with **Y** = 1 if bleeding or **Y** = 0 if no bleeding. **Y** can also be continuous, such as in LOS.

In observational studies the vector of covariates **X** could be related to both the potential outcome of interest **Y** and the treatment administered **Z**. Since **Z** and **Y** are affected by **X**, **Z** is therefore not independent of **Y**. Since **X** can affect both the probability of treatment and the probability of the outcome it is referred to as confounders. Ignoring the potential confounding effects of patient characteristics may lead to biased estimation of the treatment effect.[13] Thus, it is vital in observational studies to address potential bias due to confounding. Unbiased estimates of the parameters of interest can be obtained after controlling for observed characteristics, for example with propensity scores (PS) estimated via parametric regression methods.[14] However, unbiased estimates can only be obtained if the regression models are correctly specified. To minimize bias due to model misspecification, machine learning methods have been employed recently to find the best-fitting model for the data, thus producing consistent estimates of the propensity scores. These methods have been shown in numerous simulation studies to be able to reduce bias due to model misspecification.[15, 16]

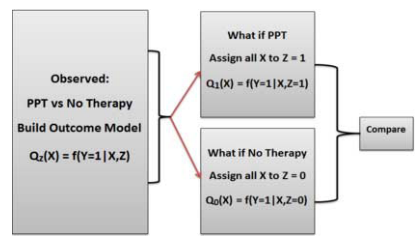


Figure 2: Potential Outcome (What If) Model

To define the parameter of interest, the problem is viewed under the potential outcome framework first introduced by Rubin [17] and the recent population intervention model proposed in Hubbard and Van Der Laan.[18] The potential outcome model defines the effect of the possible levels of treatment **Z** = 1 or **Z** = 0 for each patient and allows for the consideration and estimation of what would have happened if a patient receives a particular treatment, possibly contrary to what the patient actually received. This counterfactual procedure allows for the definition of summary measures that quantify the effect of the treatment. Figure 2 illustrates the potential outcome modeling procedure. The general procedure is to use a model to estimate the outcome of interest had the entire population been treated or not treated and compare the two estimates.

With the goal of reducing bleeding risk and optimization of plasma transfusion, this work is interested in studying for from those who received PPT, what would have happen if they did or did not receive PPT. Thus, measures like the average treatment effect for the treated (ATT) or the average treatment effect for the control (ATC) is of interest. ATC is defined as $\psi = \mathsf{E}_X[\mathsf{E}[Y \mid Z = 0, X] - \mathsf{E}[Y]$, where E_X denotes the expectation with respect to **X** (see [18, 19] for more details). ATC can be regarded as a type of attributable risk, since it compares the overall mean of the outcome to the mean of the population of interest average over strata of **X**.

Estimation Methods

Assuming no other unobserved variables are present, the relationships between the observed variables are shown graphically in Figure 1. From the graphical representation, the

factorized data likelihood can be written as:

$$\Pr(X,Y,Z) = \Pr(X)\overbrace{\Pr(Z \mid X)}^{g}\overbrace{\Pr(Y \mid X,Z)}^{Q_z}$$

Node **Z** in Figure 1 is of great interest in causal inference as the goal is to determine what happens to the outcome when some intervention is done on **Z**. The probability $g(Z \mid X) = \Pr(Z \mid X)$ represents the propensity or the causal disposition of the treatment to produce or create some outcome. $Q_z(X)$ is the outcome model which represents the mean value of the outcome **Y** given **Z** and **X**. With the factorized data likelihood, different approaches can be used to answer the question *"what and how big is the effect of PPT on PB ?"*

For a meaningful measure of treatment effect, two robust estimation methods are considered for estimating the parameter ψ: Double Robust (DR) and Targeted Maximum Likelihood (TMLE) Estimation. More in-depth treatment of these estimators can be found in [18, 19].

Doubly Robust Estimator

The Doubly Robust (DR) estimator for ATC is derived in [18, 19] and given by

$$\psi_{DR} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{I(Z_i = 0)}{\hat{g}(X_i, Z_i = 0)} - 1\right)Y_i - \left(\frac{I(Z_i = 0)}{\hat{g}(Z_i = 0 \mid X_i)} - 1\right)\hat{Q}_0(X_i)\right]$$

Where I is the indicator function, $\hat{g}(Z = 0 \mid X)$ and $\hat{Q}_0(X)$ are estimates of $g(Z = 0 \mid X)$ and $Q_0(X)$ respectively.

Doubly robustness means that ψ_{DR} is a consistent estimate of the causal treatment effect when either $\hat{g}$ or $\hat{Q}_0$ is a consistent estimate of g and Q_0 respectively.

Targeted Maximum Likelihood Estimator (TMLE)

From the definition of the DR estimator, the occurrence of g in the denominator shows that the estimator may blow up when $\hat{g}$ is not well-bounded. For some patients where estimates of g is close to zero, ψ_{DR} may be unstable. The lack of boundedness in $\hat{g}$ is a violation of one of three identifying assumptions: namely, the positivity or experimental treatment assumption (ETA) that requires $0 < g(Z \mid X) < 1$. See [19] for a more thorough discussion of these assumptions.

To solve this problem, the Targeted Maximum Likelihood Estimator (TMLE) was proposed in [20]. Like DR, TMLE is also double robust and locally efficient. The technical details of this estimator are beyond the scope of this paper and the interested reader is referred to the cited reference for more details.

Machine Learning Methods

Traditionally, the estimation of the functions g or Q_0 is performed using generalized linear models such as logistic regression or ordinary least squares. However, these parametric approaches are prone to model misspecification. An alternative and more attractive approach is to employ machine learning methods. Machine learning aims to infer the true relationship between the outcomes and covariates through a learning procedure. Bias in the estimates resulting from model misspecification can thus be reduced.

This work makes aggressive use of six machine learning methods: Support Vector Machine (SVM), Neural Networks Using Model Averaging (NNET), AdaBoost (Stochastic gradient boosting), randomForest (RF), random k-nearest neighbor (rKNN), and Generalized Boosted Regression Models (GBM). A comprehensive review of these supervised learning techniques can be found in.[21] rKNN is a recent algorithm that builds multiple k-nearest neighbor classifiers or regression models and combines them using a similar strategy as in the RF algorithm.

Experiments

This section describes the experimental setup: the study population, covariate balancing, and variable importance.

Study Population

To be considered for this study, patients must meet the following criteria: age $\geq$ 18 years, noncardiac surgery and an INR ≥ 1.5 in the 30 days preceding surgery. See Jia et al [11] for more details on the selection criteria and data source.

Between January 1, 2008, and December 31, 2011, a total of 155,492 patients aged $\geq$ 18 years underwent noncardiac surgery at the participating institution. Of them, 14,743 had an INR measured within 30 days of the index surgical procedure, with 1,234 having an INR $\geq$ 1.5. This latter group comprised the study population.

Baseline patient demographics include age, weight, gender, and American Society of Anesthesiologists Physical Status I-V classifications (ASA). Disease conditions included myocardial infarction, congestive heart failure, cerebrovascular disease, dementia, chronic pulmonary disease, connective tissue disease, diabetes mellitus, tumor, etc. Preoperative laboratory test results included INR, hemoglobin, creatinine, albumin, and activated partial thromboplastin time. Preoperative medications such as aspirin, clopidogrel, and heparin were also included in the models.

Covariate Balancing and Variable Importance

In causal inference, the models do not have to only fit the data well, but also represent (and hence balance) the features of patients who received PPT and patients who did not. Thus an initial investigation using RF was performed to determine if there are confounding covariates in the study population. See Jia et al [11] for the results from standard univariate and multivariate statistical approach in balancing the data.

First, RF variable importance measure was used to identify features that are statistically associated with PPT (results not shown). The method described in Luz Calle et al [22] was used to estimate p-values. Then a robust Covariate Balancing Propensity Score (CBPS) method [23] was applied to estimate propensity scores by maximizing the average treatment effect (ATE). The estimated propensity scores were then used in a non-parametric matching algorithm described in Ho [24] to pair patients who received PPT to patients who did not. Table

1 shows, under the covariate balancing columns, the RF permutation variable importance p-values based on the Gini impurity measure for the adjusted and unadjusted covariates. Similarly to the results in Jia et al,[11] the RF variable importance identified many between-group differences. However, after performing CBPS, the all non-significant p-values under the *Adjusted* column of Table 1 indicates that the propensity score matching was effective in addressing covariate imbalance.

Second, the same RF permutation variable importance measure was again applied to the CBPS balanced data to identify statistical relevant predictors of perioperative bleeding. The last column of Table 1 shows that Clopidogrel , Peptic Ulcer, Chronic Renal Failure, and Creatinine are significant risk factors for bleeding at the 5% significance level.

Results

This section presents the main results: Estimates of treatment effect of preoperative plasma transfusion on perioperative bleeding and other important patient outcomes. For each estimation method described in section 3, the g and Q_z functions are estimated using the six machine learning methods. For comparison with standard regression methods, g is also estimated with logistic regression (LR) while Q_z is estimated with LR or linear regression depending on whether the outcome is binary or continuous .

For each method, the parameter estimate, standard error, and significance represented by two sided p-value are reported. All computations were performed using a modified version of the R statistical package multiPIM,[25] a causal inference approach to variable importance analysis. The standard 5-fold cross-validation training procedure was followed for all experiments.

Table 1: Random forest variable importance p-values for balanced and unbalanced data and risk factors for bleeding

	Covariate Balancing		Bleeding
Variables	Unadjusted	Adjusted	Risk Factors
Clopidogrel	**0.050**	0.950	**0.030**
Peptic Ulcer	0.673	0.459	**0.033**
Chronic Renal Failure	0.741	0.094	**0.040**
Creatinine	0.864	0.218	**0.040**
PLT	1.000	0.455	0.080
ASA Recoded	**0.012**	0.589	0.237
Hemiplegia	0.553	0.898	0.263
Coumadin	0.493	0.359	0.321
Peripheral Vascular	0.770	0.838	0.435
Dementia	0.553	0.697	0.465
MI	0.387	0.790	0.466
Age	0.365	0.218	0.474
Cancer	0.864	0.721	0.488
Lymphoma	0.447	0.357	0.512
DM organ damage	0.367	0.545	0.516
Heparin	0.072	0.403	0.559
Aspirin	0.473	0.242	0.565
Hemoglobin	0.679	0.253	0.576
Cancer meta	0.870	0.647	0.602
INR	0.946	0.453	0.616
Connective Tissue Disease	0.176	0.481	0.703
Gender	0.922	0.818	0.724
Leukemia	0.926	0.425	0.796
Cerebrovascular Disease	0.645	0.896	0.806
Pulmonary Disease	0.220	0.118	0.854
Emergency	**0.002**	0.561	0.872
DM	0.906	0.737	0.920
Procedure categories	**0.008**	0.541	0.976
Congestive Heart Failure	0.904	0.828	0.982
Liver Disease	0.445	0.224	1.000

Table 2 presents estimates of ψ , the average treatment effect of PPT on PB and other outcomes for the untreated. Assuming that the ETA assumption hold from Young et al,[19] a value of $\psi = -0.05$ in the table for a binary outcome can be interpreted as: "The effect of preoperative plasma transfusion, given that all patients in the population are not transfused, is to reduce the risk of the outcome by 5 percentage points".

Table 2: Impact of plasma transfusion on perioperative bleeding and other important patient outcomes

Estimator	Outcome	Statistics	LR	SVM	NNET	AdaBoost	RF	rKNN	GBM
DR	PB	ψ	-0.010	-0.016	-0.014	-0.011	-0.008	-0.015	-0.022
		SE	0.005	0.004	0.004	0.004	0.002	0.005	0.006
		p-values	0.186	**0.000**	**0.007**	**0.032**	**0.008**	**0.006**	**0.001**
	Intra-RBC	ψ	-0.010	-0.017	-0.016	-0.011	-0.008	-0.015	-0.022
		SE	0.005	0.004	0.004	0.004	0.003	0.004	0.006
		p-values	0.186	**0.000**	**0.001**	**0.040**	**0.009**	**0.007**	**0.001**
	Re-OP	ψ	-0.014	-0.027	-0.017	-0.014	-0.013	-0.022	-0.031
		SE	0.005	0.005	0.005	0.005	0.003	0.005	0.006
		p-values	**0.049**	**0.000**	**0.002**	**0.019**	**0.001**	**0.000**	**0.000**
	ICU-LOS	ψ	-0.058	-0.301	-0.244	-0.076	0.018	-0.200	-0.171
		SE	0.121	0.089	0.129	0.113	0.070	0.094	0.118
		p-values	1.000	**0.001**	0.119	1.000	1.000	0.066	0.299
	ICU-Care	ψ	-0.014	-0.022	-0.017	-0.011	-0.012	-0.021	-0.024
		SE	0.005	0.004	0.005	0.003	0.003	0.005	0.005
		p-values	0.057	**0.000**	**0.002**	**0.003**	**0.001**	**0.000**	**0.000**
TMLE	PB	ψ	-0.010	-0.018	-0.013	-0.011	-0.012	-0.008	-0.022
		SE	0.005	0.004	0.005	0.004	0.003	0.005	0.006
		p-values	0.185	**0.000**	**0.031**	**0.041**	**0.000**	0.459	**0.001**
	Intra-RBC	ψ	-0.010	-0.017	-0.012	-0.011	-0.013	-0.017	-0.022
		SE	0.005	0.004	0.005	0.004	0.003	0.005	0.006
		p-values	0.185	**0.000**	0.076	**0.049**	**0.000**	**0.002**	**0.001**
	Re-OP	ψ	-0.014	-0.029	-0.022	-0.013	-0.016	-0.007	-0.031
		SE	0.005	0.005	0.005	0.005	0.003	0.005	0.006
		p-values	**0.047**	**0.000**	**0.000**	**0.050**	**0.000**	**1.000**	**0.000**
	ICU-LOS	ψ	-0.058	-0.157	-0.171	-0.136	-0.085	-0.195	-0.170
		SE	0.121	0.105	0.113	0.109	0.071	0.094	0.119
		p-values	1.000	0.267	0.263	0.426	0.456	0.075	0.309
	ICU-Care	ψ	-0.014	-0.077	-0.019	-0.008	-0.015	-0.019	-0.025
		SE	0.005	0.008	0.005	0.003	0.003	0.004	0.005
		p-values	0.055	**0.000**	**0.000**	0.099	**0.000**	**0.000**	**0.000**

Table 3: Accuracy measures for predicting perioperative bleeding if plasma transfusion was not administered

Classifier	PCC	PCC.se	AUC	AUC.se	sens	sens.se	spec	spec.se
AdaBoost	0.792	0.017	0.868	0.016	0.734	0.031	0.824	0.020
GBM	0.791	0.017	0.864	0.016	0.772	0.029	0.801	0.021
LR	0.786	0.017	0.860	0.016	0.734	0.031	0.816	0.020
NNET	0.767	0.018	0.852	0.016	0.717	0.031	0.796	0.020
RF	0.808	0.016	0.863	0.016	0.673	0.033	0.884	0.017
rKNN	0.765	0.018	0.842	0.017	0.709	0.032	0.796	0.021
SVM	0.739	0.018	0.823	0.018	0.734	0.030	0.743	0.022

Overall, all algorithms confirmed that PPT increases the risk of PB and all other considered outcomes. However, machine learning methods turn to generate significant results as illustrated by the small p-values. The significant results at the 5% level are highlighted. Results from machine learning methods show that PPT significantly impacts PB and Intraoperative RBC transfusion by 1-2 %, and need for ICU care by 1-7%. While PPT negative impacts ICU length of stay, hospital length of stay, and mortality (results not shown), the effects are not statistically significant.

The performance measures of the classifiers in predicting PB, assuming the counterfactual that no patient was administered PPT, is shown in Table 3. PCC is the percent of cases correctly classified (accuracy), AUC is the area under the receiver operating characteristic curve, while sens and spec are the sensitivity and specificity, respectively. Standard errors are also shown. The overall performance of the classifiers was very good with boosting methods tending to show a slightly better performance.

The performance of the two estimators (DR and TMLE) appear to be quite comparable, especially for logistic and linear regression. However, as noted earlier, DR may be unstable when the estimated propensity scores are close to zero. No such instability was observed for the analysis in this paper.

Conclusion and Future Work

This work employed advanced machine learning methods to measure the effectiveness of plasma transfusion from observational data. The method applies a causal modeling framework that establishes a causal relationship between administering plasma transfusion and perioperative bleeding based on two robust estimation procedures. The paper provides meaningful interpretation of the estimation results which are more intuitive to understand than estimated coefficients from regression models.

Results from the analysis show that population wise, the action of not administering plasma product has the impact of reducing bleeding and positive effects on other patient important outcomes. These results are not new, as several authors have derived these results. However this paper takes a causal approach based on modern machine learning to quantify in a meaningful way the effect of not transfusing. Given that, on average, no plasma transfusion reduces the risk of bleeding, an interesting problem warranting further investigation is: "Can machine learning methods be used to identify a sub-population for which no plasma transfusion increases the risk of bleeding?" This paper is an initial step to answering this important question.

Acknowledgments

This research was supported in part by funding from Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery.

References

[1] Paolo Feltracco, Maria Luisa Brezzi, Stefania Barbieri, et al. Blood loss, predictors of bleeding, transfusion practice and strategies of blood cell salvaging during liver transplantation. World journal of hepatology, 5(1):1, 2013.

[2] Fengyu Zheng, Frank P Cammisa Jr, Harvinder S Sandhu, et al. Factors predicting hospital stay, operative time, blood loss, and transfusion in patients undergoing revision posterior lumbar spine decompression, fusion, and segmental instrumentation. Spine, 27(8):818–824, 2002.

[3] Anne F Eder, Ross M Herron Jr, Annie Strupp, et al.. Effective reduction of transfusion-related acute lung injury risk with male-predominant plasma strategy in the american red cross (2006-2008). Transfusion, 50(8):1732–1742, 2010.

[4] US Department of Health, Human Services, et al. The 2009 national blood collection and utilization survey report. Washington, DC: US Department of Health and Human Services, Office of the Assistant Secretary for Health, 2011.

[5] Yves Ozier, Fabienne Pessione, Emmanuel Samain, Françoise Courtois, French Study Group on Blood Transfusion in Liver Transplantation. Institutional variability in transfusion practice for liver transplantation. Anesthesia & Analgesia, 97(3):671–679, 2003.

[6] N Shehata, G Naglie, AA Alghamdi, et al. Risk factors for red cell transfusion in adults undergoing coronary artery bypass surgery: a systematic review. Vox sanguinis, 93(1):1–11, 2007.

[7] Walter Dzik and Arjun Rao. Why do physicians request fresh frozen plasma? Transfusion, 44(9):1393–1394, 2004.

[8] Omar I Abdel-Wahab, Brian Healy, and Walter H Dzik. Effect of fresh-frozen plasma transfusion on prothrombin time and bleeding in patients with mild coagulation abnormalities. Transfusion, 46(8):1279–1285, 2006.

[9] Lorne L Holland, Tisha M Foster, Richard A Marlar, and Jay P Brooks. Fresh frozen plasma is ineffective for correcting minimally elevated international normalized ratios. Transfusion, 45(7):1234–1235, 2005.

[10] Guangxi Li, Sonal Rachmale, Marija Kojicic, et al. Incidence and transfusion risk factors for transfusion-associated circulatory overload among medical intensive care unit patients. Transfusion, 51(2):338–343, 2011.

[11] Qing Jia, Michael J. Brown, Leanne Clifford, et al. Preoperative plasma transfusion does not decrease the risk of bleeding in surgical patients who have abnormal coagulation tests. Submitted to Lancet 2014.

[12] Mary E Northridge. Public health methods–attributable risk as a link between causality and public health action. American journal of public health, 85(9):1202–1204, 1995.

[13] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1):41–55, 1983.

[14] Marshall M Joffe and Paul R Rosenbaum. Invited commentary: propensity scores. American journal of epidemiology, 150(4):327–333, 1999.

[15] Daniel Westreich, Justin Lessler, and Michele Jonsson Funk. Propensity score estimation: neural networks, support vector machines, decision trees (cart), and meta-classifiers as alternatives to logistic regression. Journal of clinical epidemiology, 63(8):826–833, 2010.

[16] Brian K Lee, Justin Lessler, and Elizabeth A Stuart. Improving propensity score weighting using machine learning. Statistics in medicine, 29(3):337–346, 2010.

[17] Rubin, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology 66.5 (1974): 688.

[18] Alan E Hubbard and Mark J Van Der Laan. Population intervention models in causal inference. Biometrika, 95(1):35–47, 2008.

[19] Jessica G Young, Alan E Hubbard, B Eskenazi, and Nicholas P Jewell. A machine-learning algorithm for estimating and ranking the impact of environmental risk factors in exploratory epidemiological studies. 2009.

[20] Mark J Van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer, 2011.

[21] C Crisci, B Ghattas, and G Perera. A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling, 240:113–122, 2012.

[22] M Luz Calle, Victor Urrea, A-L Boulesteix, and Nuria Malats. Auc-rf: A new strategy for genomic profiling with random forest. Human heredity, 72(2):121–132, 2011.

[23] Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1):243–263, 2014.

[24] Daniel E Ho. MatchIt: nonparametric preprocessing for parametric causal inference. PhD thesis, Departments of Mental Health and Biostatistics, Johns Hopkins Bloomberg School of Public Health, 1737.

[25] Stephan J Ritter, Nicholas P Jewell, and Alan E Hubbard. R package multipim: A causal inference approach to variable importance analysis. Journal of Statistical Software, 57(8), 2014.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-726

On the Automated Segmentation of Epicardial and Mediastinal Cardiac Adipose Tissues Using Classification Algorithms

Érick Oliveira Rodrigues[a], Felipe Fernandes Cordeiro de Morais[b], Aura Conci[a]

[a] *Institute of Computing, Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil*
[b] *School of Medicine, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil*

Abstract

The quantification of fat depots on the surroundings of the heart is an accurate procedure for evaluating health risk factors correlated with several diseases. However, this type of evaluation is not widely employed in clinical practice due to the required human workload. This work proposes a novel technique for the automatic segmentation of cardiac fat pads. The technique is based on applying classification algorithms to the segmentation of cardiac CT images. Furthermore, we extensively evaluate the performance of several algorithms on this task and discuss which provided better predictive models. Experimental results have shown that the mean accuracy for the classification of epicardial and mediastinal fats has been 98.4% with a mean true positive rate of 96.2%. On average, the Dice similarity index, regarding the segmented patients and the ground truth, was equal to 96.8%. Therfore, our technique has achieved the most accurate results for the automatic segmentation of cardiac fats, to date.

Keywords:

Segmentation; Classification; Epicardial fat; Mediastinal fat; Adipose tissue; Computed tomography; Data mining; Cardiac fat.

Introduction

Extensive studies address the importance of the epicardial and mediastinal (due to a nomenclature inconsistency some call it pericardial) fats and their correlation with pathogenic profiles, risk factors and diseases [1,2]. Some work [3] associates both mediastinal and epicardial fats to carotid stiffness; others [4] associate them to atherosclerosis, coronary artery calcification and others health risk factors. Sicari *et al.* [5] have shown as well how mediastinal fat rather than epicardial fat is a cardiometabolic risk marker.

An increasing demand for medical diagnosis support systems has been observed jointly to the computational evolution in the last years. Systems of this kind speed up the tedious and meticulous analysis conducted by physicians or technicians on patients' medical data. In several cases there is a huge amount of data to be analyzed and, therefore, the diagnosis may lack of precision and suffer noticeable inter and intra-observer variation [6].

The automated quantitative analysis of epicardial and mediastinal fats certainly adds a prognostic value to cardiac CT trials with an improvement on its cost-effectiveness. Iacobellis *et al.* [7] have shown that the epicardial fat thickness and coronary artery disease correlate independently to obesity, which supports individual segmentation of these adipose tissues rather than merely and simply estimating its volume based on the patient overall fat.

Latterly, three imaging techniques appear suitable for quantification of these adipose tissues, namely Magnetic Resonance Imaging (MRI), Echocardiography and Computed Tomography (CT). Each has been used in several medical studies [5,8,9]. However, CT provides a more accurate evaluation of fat tissues due to its higher spatial resolution compared to ultrasound and MRI [10]. In addition, CT is also widely used for evaluating coronary calcium score [9].

Literature Review

Semi-automated segmentation methods for epicardial fat have been proposed since 2008. Dey *et al.* [11], for instance, apply a preprocessing step to remove all other structures from the heart by using a region growing strategy. An experienced user is required to scroll through the slices and to place from 5 to 7 control points along the pericardium border if visible. Thus, Catmull-Rom cubic spline functions are automatically generated to obtain a smooth closed pericardial contour. Finally, since the epicardial fat is inside this contour it is simply accounted by thresholding. In [12], a method is proposed for the segmentation of abdominal adipose tissue. The work of Kakadiaris *et al.* [13] has further extended the method to the segmentation of the epicardial fat.

Coppini *et al.* [10] focused on reducing the user intervention. On their method, an expert is still necessary to scroll through the slices between the atrioventricular sulcus and the apex in order to place some control points on the pericardium layer. The amount of required control points is not clearly described. Nevertheless, the amount of slices to be analyzed is apparently less than that of the method proposed by Dey *et al.* [11]. Coppini et al. also present their solution in a 3D space. The overall focus of their work was to describe their method mathematically. Nevertheless, the authors did not address the achieved general accuracy.

Barbosa *et al.* [14] proposed a more automated segmentation method for the epicardial fat. They use the same preprocessing method from Dey *et al.* [11] and further apply a high level step for identification of the pericardium by tracing lines originating from the heart's centroid to the pericardium layer and interpolating them with a spline. Although this approach may be of simple complexity and highly applicable for virtually any proposed method in the field, the reported results are not impressive. Only 4 out of 40 images were correctly segmented in a fully automatic way.

As far as we know, Shahzad *et al.* [15] proposed the first fully automated method for epicardial fat segmentation. Their method uses a multi-atlas based approach to segment the pericardium. The multi-atlas approach is based on registering several atlases (8 in this case) to a target patient and fusing these transformations to obtain the final result. They selected

98 patients for evaluation and reported a Dice similarity index of 89.15% to the ground truth and a low rate of approximately 3% of unsuccessful segmentations. Nevertheless, they did not provide any measurements of the overall processing time.

Ding *et al.* [16] proposed, in 2010, an approach similar to the method of Shahzad *et al.* [15]. They segment the pericardium using an atlas approach, which consists of a minimization of errors after applying transformations to the atlas along with an active contour approach. The mean Dice similarity coefficient was 93% and they claim that their result was achieved in 60 seconds on a simple personal computer. Although their segmentation seems to be the most precise in the literature, the reported computational time is oddly described. That is, 60 seconds is considered too fast for segmenting and transforming an entire patient scan, which consists of roughly 50 images. They also present a work [17] that segmented the aorta instead of the pericardium and compare their time (60 seconds) to the 15 minutes of the former. If these 60 seconds correspond to just the time it takes for the algorithm to minimize the transformations, then this comparison is not feasible. Furthermore, they report that their approach had the atlases' images pre-aligned to a standard orientation and that there is a comparison with just one of these atlases to speed up the process. The remaining pericardium contour follows the pre-aligned pattern, which is a reported limitation. Besides, they did not describe how each one of these atlases is chosen as suitable for each possible case.

Materials and Methods

We define the fat located within the epicardium as epicardial fat, corroborating with the majority of published works [4-8,10]. Furthermore, by following the same "first outer anatomical container" logic, we conclude that mediastinal fat is the best definition for the fat located on the external surface of the heart or fibrous pericardium. In other words, the mediastinal fat is located within the mediastinal space as long as it is not epicardial (i.e., it is not located within the epicardium). Furthermore, we have used CT scans from two manufactures (Siemens and Philips), which configures this work as being multi-manufacturer.

Molteni [18] associates values around -100 Hounsfield Units (HU) to the overall fat of the human body. Coppini *et al.* [10] and Shmilovich et al. [19] defined the cardiac adipose tissue range as starting from -190 to -30, Spearman *et al.* [20] defined it as from -195 up to -45, and Shahzad *et al.* [15] defined it as from -200 to -30. In this work, we consider the largest addressed interval for the cardiac fat, which corresponds to the one considered in the work of Shahzad *et al.*, *i*.e., from -200 to -30 HU.

Overview

We propose an automatic segmentation for the epicardial and mediastinal fats based on two main principles, namely (1) an intersubject registration and (2) a classification step. For the whole segmentation process we have used CT images ranging from -200 to -30 HU on the axial plane. However, we believe that our methodology can be easily adapted to other ranges and also to other modalities.

Image registration can be defined as the process of matching characteristics of images in order to find alignments that minimize the variation between overlapping pixels [22]. Such processes are included in panoramic assemblages, medical images, time series alignments [23,24] and many others. Registration is also alternatively treated as an optimization problem with the goal of finding the spatial mapping that will bring images, parts of them, or even a combination of these parts into minimal variation.

Machine learning algorithms are often divided in two main categories: (1) supervised and (2) unsupervised methods. The algorithm is categorized as supervised when it explicitly evaluates the class attribute of a training set as the predictive label desired to attach to an incoming unlabeled instance. Furthermore, when this assumption is formalized, the class attribute heavily induces the generated predictive model. However, without the formalization, the algorithm is defined as unsupervised and the class induces no heavy influence but of a normal attribute to the predictive model when, of course, it is not disregarded from the training pace. Classification algorithms are always categorized as supervised learning methods while clustering algorithms are often unsupervised.

Registration

In a previous work we extensively described the methodology of our intersubject registration [25]. Summing up, it is composed of (1) creating an atlas of the retrosternal area, (2) using this atlas to recognize the same area of an incoming instance, and (3) using a heuristical method to reinforce the chosen position. The atlas assemblage is done by (1) converting a number of CT slices to the fat range where black represents the background, (2) manually selecting the position of the retrosternal area, (3) thresholding each selected area to 0 or 1 (binary image) and finally, (4) performing an arithmetic mean of the binary images. The final image of Figure 1 is the result of the described process conducted for 10 randomly chosen patients. Thereafter, the proposed registration consists of fusing two approaches: a landmark with an atlas approach. The landmark approach is defined as when two selected points from the subjects to be registered are used as reference (in this case, the center of the retrosternal area) [26]. Finally, the retrosternal area is defined as the region at the back of the sternum and, in this case, it frequently captures parts of the heart if it is relatively close to the thorax as well as few parts of the thorax as shown in Figure 1 by the rectangles in red.

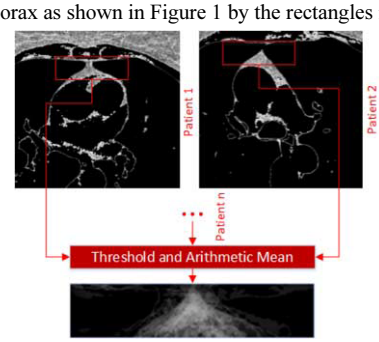


Figure 1 – Atlas of the retrosternal area.

Once the atlas is assembled it is considered immutable for any subject to be registered. The following step is to displace the atlas image on top of any cardiac CT image that is on the proper fat range and that has the retrosternal area visible and, to compute a similarity score associated to every possible position of the atlas. A weighted mutual information (WMI), as shown in Equation (1), was chosen for computing the similarity score. Hence, we are autonomously recognizing the pattern of the retrosternal area on an arbitrary patient in order to suppose its location. After assuming its location on the basis of the WMI score, a heuristical confirmation method was applied to reinforce the chosen position. Once the position is set, all the retrosternal areas of the patients being registered are aligned, based on the recognized landmark, to a common

position. It is important to emphasize that this registration step is required to be done on a single slice of each patient; and the same transformation is applied to the remaining.

$$WMI_{y,x}(F,M,g) = \left(\sum_{f\in F}\sum_{m\in M}\frac{1}{|f-m|+1}\rho_{FM}(f,m)\log_g\frac{\rho_{FM}(f,m)}{\rho_F(f)\rho_M(m)}\right) \quad (1)$$

The variable F represents the fixed slice, M represents the moving atlas and g stands for the base of the logarithm. In the traditional formulation of the mutual information, each event or object specified by (f,m) is weighted by the corresponding joint probability $\rho_{FM}(f,m)$. This assumes that all objects or events are equivalent apart from their probability of occurrence. However, in some applications, it may be the case that certain objects or events are more significant than others, or that certain types of associations are more semantically important than others. Thus, we combined the mean difference measure with the mutual information, originating a weighted measure by the difference mutual information measure previously shown in Equation (1).

Classification

The classified segmentation can be viewed as a simple iteration through a set of pixels of an image (or voxels of a 3D model) where a set of features (i.e., characteristics) related to the iterated pixel, voxel, or surrounding area is extracted. The set of features is commonly called the features vector. The vectors are often united to compose a dataset that is provided as input to a classification algorithm.

In order to generate a concise predictive model we need to provide reliable data for training the classification algorithm. Therefore, two specialists, one being a physician and the other being a computer scientist, have manually segmented the epicardial and mediastinal adipose tissues of 20 patients (10 male and 10 female). Thus, our ground truth contains approximately 900 manually segmented cardiac CT images. It is important to highlight that prior to the manual segmentation, the images were already registered by our methodology. The created ground truth is available at [27]. The black value (0) represents the background and these pixels are excluded from the feature extraction and from the statistics. All other pixel values represent the selected cardiac fat range: (-200,-30) HU.

Similarly to the approach of Rikxoort *et al.* [28] for the classified segmentation of the liver, we have selected features: the pixel grey level and the position x, y and z, where z is the index of the slice. Besides, we have also selected the x and y positions relative to the center of gravity of the image and texture-based features from a neighborhood of variable size that encapsulates the iterated pixel at its center, i.e., a surrounding window of pixel values. Some features were selected on a theoretical basis and extracted from this neighborhood, such as: (1) a simple arithmetic mean of the grey levels, (2) moments of the co-occurrence matrix, (3) geometric moments, (4) run percentage, (5) grey level non-uniformity and (6) a 1D Gaussian-weighted mean of the grey values [24]. There are three possible classes for a pixel in our problem, namely, (1) epicardial fat, (2) mediastinal fat or (3) pixel of the pericardium (or unknown). These classes were evaluated separately and, therefore, the problem was reduced to a binary classification (true or false) for each one of these three possible classes [24]. If an incoming pixel is classified as epicardial and mediastinal, the result is a hybrid class of both. If it is classified as pericardium and has no current label, the result is also the same hybrid (further represented in yellow in Figure 3).

Results

Following feature selection, some classifiers were selected for a first evaluation of the convergence speed. For this trial, we extracted the features from the slices of just one single patient (approximately 13 107 200 features vectors) in order to speed up the analysis. All the classification and clustering algorithms in Weka [30] were selected for evaluation, some of these algorithms are, namely, the Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), Naïve Bayes, Radial Basis Function (RBF) Network, Random Trees, C4.5 (or J48), Primal Estimated Sub-Gradient Solver for SVM (SPegasos), REPTree, k-Nearest Neighboor (kNN), Multilayer Perceptron and others. The parameters used by the evaluated classifiers were based on their standards with a small tweaking in some cases and the best result was selected. The algorithms present in Table 1 were the only ones that converged within an interval of 200 seconds (constructing and evaluating the predictive model). The values stand for mean accuracies and times of the three possible classes on a neighborhood size of 5x5 pixels.

Table 1 – Accuracies and convergence time on a single patient

Algorithm	Accuracy	Time (s)	Acc/Time
J48Graft	99.0%	132.86	0.75
RandomForest	98.9%	112.57	0.88
REPTree	98.9%	10.34	9.56
J48	98.9%	151.23	0.65
SimpleCart	98.9%	108.78	0.91
SMO	98.3%	58.66	1.68
RandomTree	97.5%	8.0	12.19
RBFNetwork	96.8%	3.48	27.82
SPegasos	96.8%	15.77	6.14
DecisionStump	96.8%	52.34	1.85
HyperPipes	94.8%	0.04	2370.0
NaiveBayes	86.0%	55.48	1.55

At first glance, the REPTree algorithm appears to be the best choice since it achieved a great accuracy in a relatively short time, as compared to the others. Also of high interest was the HyperPipes result. In this case, it returned a good accuracy almost instantaneously. However, when applied to a bigger dataset and a set of images of distinct patients, the results of the HyperPipes algorithm happen to differ drastically.

Overall Evaluation

To avoid unfair comparisons of classifiers we have further evaluated some of the algorithms of Table 1 along with a variation on the neighborhood size (modified during the extraction of features). Hypothetically, one may consider that some classifiers perform better on a neighborhood of a certain specific size. In Figure 2, a chart that represents the accuracy (vertical axis) of each evaluated classifier (horizontal axis) versus the variation of the neighborhood size in pixels (orthogonal to the previous axes) is shown. The assessed CT images were 512x512 pixels wide.

The accuracies in Figure 2 were achieved using 66% random-selected split method as test mode from the data of 16 patients. The 66% split test mode is defined as randomly selecting 66% of the instances for training and using the remaining for evaluation of the predictive model. The reason for choosing the split method and 16 patients was to reduce the huge convergence time due to the great amount of data and, consequencely to speed up the analysis. The extracted features composed datasets are approximately 1.5 gigabytes for each neighborhood size that was provided to each classifier. The

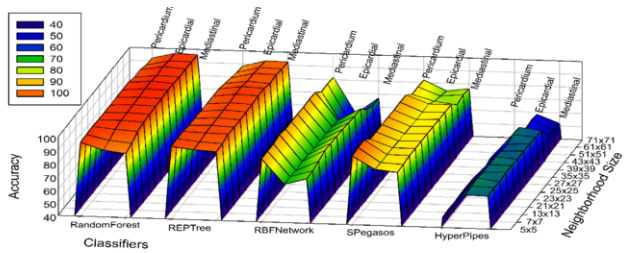


Figure 2 – Accuracies of selected classifiers versus pixels neighorboods of distinct sizes.

period to train and evaluate the predictive model was, in some cases, up to 20 minutes for each possible combination of neighborhood size and classification algorithm.

The five algorithms shown in Figure 2 were the quickest on this large dataset (16 patients). The HyperPipes was the fastest and always converged virtually within 0.5 seconds but, in this case, the best accuracy that it could achieve was around 70%. We can state by this evidence that this algorithm is not generalizable. In other words, the algorithm overfitted to a single patient and failed to generalize the predictive model to a class instead of a single patient. It is a rather simple algorithm and, in fact, the achieved low accuracy was somehow supposed to happen, even in the convergence evaluation trial.

Due to the convergence time issue, we were not able to extensively assess all the possible sizes for the neighborhood in consideration for all the algorithms shown in Table 1. The REPTree algorithm did not converge remarkably faster than the remaining decision tree algorithms evaluated on this large dataset. The RBFNetwork was faster than RandomForest and SPegasos was slower but both achieved lower accuracies. SPegasos was the slowest among the algorithms in Figure 2. RandomTree and DecisionStump were just a little faster than RandomForest and REPTree but the accuracies were significantly lower; therefore, they were disregarded in the second evaluation due to the massive presence of decision tree algorithms. The J48Graft returned similar accuracies if compared to RandomForest but its convergence was approximately 1.4 times slower and, so it was impracticable to evaluate all the neighborhood sizes for this classifier. The SimpleCart, Naïve Bayes, J48 and SMO took more time to converge on the large dataset than SPegasos, therefore, they could not be precisely evaluated in the second trial.

Visualization

Figure 3 compares a single manually segmented slice (left) to the result of the proposed automatic segmentation (right). Green denotes the mediastinal fat, red represents the epicardial and blue corresponds to the pericardium. All the colored pixels represent pixels within the fat range for a CT image, therefore, there are some discontinuities on the images. Figure 4 corresponds to the same patient shown in Figure 3 but reconstructed in a 3D model. It is possible to distinguish the contour of the heart based on the epicardial fat (red color).

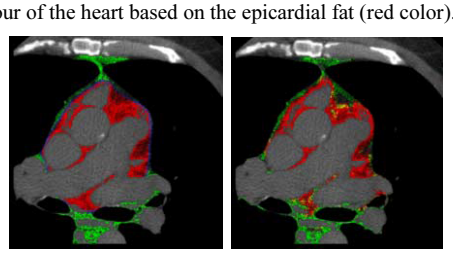


Figure 3 – Manually (left) and automatically segmented slice (right).

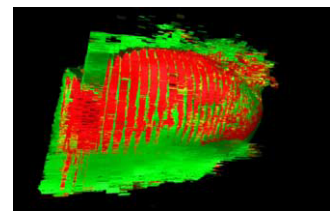


Figure 4 – 3D model of the automatically segmented patient.

Table 2 compares the results obtained by the proposed methodology to the four main related works. Barbosa *et al.* [14] and Kakadiaris *et al.* [13] use semi-automated methods, while Shahzad *et al.* [15] and Ding *et al.* [16] use fully automatic methods. The first column of the table indicates the rate of successful automatic segmentations, which are usually observed by a physician. The second column is the Dice similarity index and the third is the true positive rate. All these four works only segment the epicardial fat, therefore, we only compare our epicardial fat segmentation rates.

Table 2 – Comparison of results

Authors	Successful	Dice	T.P.
Barbosa *et al.*	10% (4/40)	-	-
Kakadiaris *et al.*	-	-	85.6%
Shahzad *et al.*	96.9% (95/98)	89.15%	-
Ding *et al.*	-	93.0%	-
This work (epicardial)	100% (82/82)	97.9%	98.3%

Conclusion

The proposed automated method achieved the best results with no need for placement of any control point, the mean accuracy for both the epicardial and mediastinal fats was 98.4% with a mean true positive rate of 96.2%. We are the first to propose an automated segmentation of the mediastinal fat and a unified methodology for the automated segmentation of both types of fat. Despite the processing time issues of our approach, this is still feasible for real time segmentation if properly adjusted. Although it requires approximately one day to segment a single patient, some heuristics could be used to speed up the classification step while an extensive selection and evaluation of features could grant an overall speed improvement. The speed up gain is also strongly related to the meticulousness of the approach. If a faster segmentation is desired, a per-area classification instead of a per-pixel classification may be applied and it should return worse but still adequate results.

Ensemble methods use multiple learning algorithms to obtain better predictive performance. In this work, we have selected a single classification algorithm (Random Forest) to segment the patients. As a further improvement, the RandomForest could be combined to the J48Graft and, perhaps, to the REPTree algorithms to increase the accuracy of the predictive

model. These three algorithms, RandomForest, J48Graft, and REPTree, were the best in our performance analysis.

Acknowledgements

E. O. Rodrigues wants to thank CAPES. A. Conci wants to thank the CNPq project 302298/2012-6. F. Morais wants to thank Ilan Gottlieb.

References

[1] Ahmad I, Hua B, Sockolow JA, et al. Correlation of Pericardial and Mediastinal Fat With Coronary Artery Disease, Metabolic Syndrome, and Cardiac Risk Factors. *Journal of Cardiovascular Magnetic Resonance*, 2009, 53(10): A283.

[2] Perseghin G, Molon G, Canali G, et al., Nonalcoholic Fatty Liver Disease Is Associated With Left Ventricular Diastolic Dysfunction in Patients With Type 2 Diabetes. *Cardiovascular and Metabolic Risk.* 2013, 58: S542-S543.

[3] Brinkley TE, Hsu FC, Ding J, et al, Pericardial fat is associated with carotid stiffness in the Multi-Ethnic Study of Atherosclerosis. *NMCD. Nutrition Metabolism and Cardiovascular Diseases*. 2011, 21(5): 332-338.

[4] de Vos AM, Stella PR, Gorter PM, et al. Relation of Epicardial and Pericoronary Fat to Coronary Atherosclerosis and Coronary Artery Calcium in Patients Undergoing Coronary Angiography. *American journal of cardiology*, 2008,102(4):380-385.

[5] Sicari R, Maria A, Petz R, et al. Pericardial Rather Than Epicardial Fat is a Cardiometabolic Risk Marker: An MRI vs Echo Study. *Journal of the American Society of Echocardiography*, 2011, 24, 1156-1162.

[6] Aspen Publishers Inc. Web-based diagnosis support system seeks to reduce delays in medical diagnoses. *Managed care outlook*. 2010, 23(13): 8.

[7] Iacobellis G, Lonn E, Lamy A, et al. Epicardial fat thickness and coronary artery disease correlate independently of obesity. *International Journal of Cardiology*. 2011, 452-454.

[8] Iacobellis G, Ribaudo MC, Assael F, et al. Echocardiographic epicardial adipose tissue is related to anthropometric and clinical parameters of metabolic syndrome: A new indicator of cardiovascular risk. *The Journal of clinical endocrinology and metabolism*. 2003, 88:5163-5168.

[9] McClelland RL, Jorgensen NW, Bild DE, et al. Coronary Artery Calcium Score and Risk Classification for Coronary Heart Disease Prediction. *Journal of the American Medical Association*. 2010, 303(16): 1610-1616.

[10]Coppini G, Favilla R, Marraccini P. Quantification of Epicardial Fat by Cardiac CT Imaging. *The Open Medical Informatics Journal*. 2010, 4: 126-135.

[11]Suzuki Y, Suzuki S, Dey D, et al. Automated quantitation of pericardiac fat from noncontrast CT. *Investigative Radiology*, 2008, 43(2): 145-153.

[12] Pednekar A. Automatic Segmentation of Abdominal Fat from CT Data. *Application of Computer Vision*, 2005, 308-315.

[13] Kakadiaris I, Bandekar A, Mao SS, et al. Automated pericardial fat quantification in CT data. *Journal of the American College of Cardiology*, 2006, 47 (4): 264A.

[14]Barbosa JG, Figueiredo B, Bittencourt N, et al. Towards automatic quantification of the epicardial fat in non-contrasted CT images. *Computer Methods in Biomechanics and Biomedical Engineering*, 2011, 14(10): 905-914.

[15]Shahzad R, Bos D, Rossi A, et al. Automatic quantification of epicardial fat volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach. *Medical Physics*, 2013, 40(9).

[16]Ding X, Terzopoulos D, Piotr J, et al. Automated Epicardial Fat Volume Quantification from Non-Contrast CT. *Medical Imaging*, 2014, 9034.

[17]Isgun I. Multi-Atlas-Based Segmentation With Local Decision Fusion—Application to Cardiac and Aortic Segmentation in CT Scans. *IEEE Transactions on Medical Imaging*, 2009, 28(7): 1000-1010.

[18]Molteni R. Prospects and challenges of rendering tissue density in Hounsfield units for cone beam computed tomography. 2013, 116 (1): 105-119.

[19]Dey D, Shmilovich H, Slomka PJ, et al. Threshold for the upper normal limit of indexed epicardial fat volume: Derivation in a healthy population and validation in an outcome-based study. *The American Journal of Cardiology*, 2011, 108(11): 1680-1685.

[20]Spearman JV, Costello P, Geyer LL, et al. Automated Quantification of Epicardial Adipose Tissue Using CT Angiography: Evaluation of a Prototype Software. *European Radiology*, 2014, 24(2): 519-526.

[21]Huang J, Hsieh MH, Chen YJ, et al. Extremely high coronary artery calcium score is associated with a high cancer incidence. *International Journal of Cardiology*, 2012, 155: 474-475.

[22]Ahmad I, Hua B, Sockolow JA. et al. Correlation of Pericardial and Mediastinal Fat With Coronary Artery Disease, Metabolic Syndrome, and Cardiac Risk Factors. *Journal of Cardiovascular Magnetic Resonance*, 2009, 53(10): A283-A283.

[23]Gerseghin G, Molon G, Canali G, et al. Nonalcoholic Fatty Liver Disease Is Associated With Left Ventricular Diastolic Dysfunction in Patients With Type 2 Diabetes. *Cardiovascular and Metabolic Risk*, 2013, 58: S542-S543.

[24]Rodrigues ÉO. Automated segmentation of epicardial and mediastinal fats using intersubject registration and classification algorithms. Thesis, Computer Science Institute, UFF, Niterói, RJ, 2015.

[25]Viergever MA, Maintz JBA. A survey of medical image registration. *Medical Ima ge Analysis*, 1998, 2(1): 1-36.

[26]Pocock AC. *Feature Selection Via Joint Likelihood*. Manchester, 2012.

[27]Groundtruth for epicardial and mediastinal fats. *Visual Lab*, 2014. http://visual.ic.uff.br/en/cardio/ctfat/

[28]Rikxoort E, Arzhaeva Y, Ginneken B. Automatic segmentation of the liver in computed tomography scans with voxel classification and atlas matching. *3D Segmentation in The Clinic: A Grand Challenge*, 2007, 101-108.

[29]Rodrigues ÉO, Conci A, Borchartt TB, et al. Comparing results of thermographic images based diagnosis for breast diseases. Proceddings of Systems, Signals and Image Processing (IWSSIP), 2014, 39-42.

[30]Frank E, G. Holmes G, M. Hall et al. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 2009, 11(1): 10-18.

Corresponding Author

Name and contact information of corresponding author.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-731

Fusing Heterogeneous Data for Alzheimer's Disease Classification

Parvathy Sudhir Pillai[a], Tze-Yun Leong[a,b], Alzheimer's Disease Neuroimaging Initiative

[a]*Medical Computing Laboratory, School of Computing, National University of Singapore, Singapore*
[b]*School of Information Systems, Singapore Management University, Singapore*

Abstract

In multi-view learning, multimodal representations of a real world object or situation are integrated to learn its overall picture. Feature sets from distinct data sources carry different, yet complementary, information which, if analysed together, usually yield better insights and more accurate results. Neuro-degenerative disorders such as dementia are characterized by changes in multiple biomarkers. This work combines the features from neuroimaging and cerebrospinal fluid studies to distinguish Alzheimer's disease patients from healthy subjects. We apply statistical data fusion techniques on 101 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We examine whether fusion of biomarkers helps to improve diagnostic accuracy and how the methods compare against each other for this problem. Our results indicate that multimodal data fusion improves classification accuracy.

Keywords:

Multimodal;Data fusion; Heterogeneous; Alzheimer's disease.

Introduction

Multimodal data fusion refers to the fusion of multiple data sources, their associated features, and (or) intermediate decisions to perform an analysis task [1]. This multimodel method has found widespread use in areas such as multimedia and sensor analyses to integrate views obtained from audio and video signals, texts and images, and others. Recent studies in medical informatics have benefitted from combining multiple data sources to better understand disease processes. In this paper, we study the impact of multimodal data fusion on classifying Alzheimers' Disease (AD) patients.

Dementia is a spectrum of neuro-degenerative disorders that lead to memory and cognitive decline, severe enough to disable a person to perform activities of daily living. AD, the most common subtype, affects close to 75% of the demented population. As of 2010, there are around 36 million affected individuals worldwide, and an enormous amount is spent on their care [2]. No definitive prevention methods/cures are available for AD. Hence, we need efficient methods to screen and study the disease early on, so that timely interventions may delay its progression.

Dementia severity is assessed by psychometric tests like Mini Mental State Examination (MMSE) and Clinical Dementia Rating (CDR), neuroimaging, protein and genomic tests, and others. Biomarkers acquired from these tests provide indicators about a person's state. The sensitivity of biomarkers varies over the stages from normal aging through Mild Cognitive Impairment (MCI) to Dementia, as evident from Figure 1 [3]. Recently, pattern classification methods have been applied to analyze these biomarkers in combinations [10, 11, 12], as the information from different biomarkers is complementary in nature. While structural Magnetic Resonance Imaging (s-MRI) has good spatial resolution to identify atrophied brain regions, functional imaging such as Fluodeoxyglucose Positron Emission Tomography (FDG-PET) reveals hypometabolism in the affected brain areas. Protein studies of the Cerebrospinal Fluid (CSF) indicate the presence of beta amyloid (Aβ42) and tau (τ) proteins which form plaques and tangles in the brain, characteristic of AD. Combining multiple related data sources yields a fused representation of the object under study. Analysing this representation yields a comprehensive picture that benefits from the interplay of statistical dependences of the data sources. Further, the analysis reduces noise in the data by averaging it out over the independent data sources.

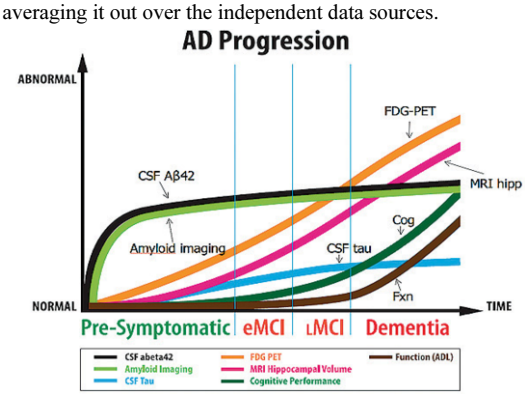


Figure 1. Biomarker sensitivity to Dementia related changes in the human brain across stages. Used with permission from the website of National Institute of Aging [3].

Motivated by these facts, we examine the effectiveness of statistical methods for fusing biomarker data to distinguish AD patients from healthy subjects (HS). On a subset of data from ADNI, we compare three data fusion methods based on:

1. Canonical Correlation Analysis (CCA)
2. Multiple Kernel Learning (MKL)
3. Collective Matrix Factorization (CMF)

While CCA ensures that the fused representation has maximally correlated features, MKL learns the optimal way to combine the features to yield the best classification accuracy. CMF is a comparatively recent method that jointly factorizes matrices that share a common dimension. We explain, implement, and test these methods on the ADNI data to compare their accuracies of classification over unimodal and prior multimodal studies.

Related Work

Quantitative fusion of medical data is very challenging because of the heterogeneity of the modalities. Two main approaches exist for combining heterogeneous information. The first approach, known as early fusion, aggregates data at the feature level into a single representation before analysis. Kernel space combination proposed by Lanckriet et al. to combine amino acid sequences and gene expressions [4], vector concatenation of Principal Component Analysis (PCA) reduced features used by Lee et al. to fuse mass spectrometry and histology information [5], and Artificial Neural Networks (ANN) used by Baez et al. to integrate various neuropsychological test scores [6] all fall under this category. Though these methods preserve inter-source dependencies, they suffer from the curse of dimensionality and hence require a large amount of training data to learn a relevant model. The second approach, known as late fusion, combines decisions from models learnt on the individual feature spaces. Various rules such as weighted combination [7], majority voting [8], likelihood maximization [9] of the decision variables have been proposed. As the fusion is at the level of decisions, there are no concerns with the dimensionality of the data. However, these methods fail to retain inter-source dependencies.

Multimodal assessments of AD and MCI were found to classify diseased individuals more accurately than unimodal methods. Zhang et al. combined MRI, PET and CSF biomarkers using multiple kernels and a coarse grid search to find the optimal kernel combination on a Support Vector Machine (SVM) classifier [10]. As compared to this discriminative approach which models the conditional distribution of variables for predicting the class labels from features, Young et al. used a variation of kernel combination with a generative Gaussian Process (GP) classifier [11]. This generative approach models the joint distribution of variables and uses likelihood maximization to learn the optimal parameters; it is shown to perform on par with the earlier discriminative approach. Gray et al. applied Random Forest (RF) proximity measures to combine MRI and PET features [12]. Though these methods provide good classification accuracy, they cannot in general support understanding of the data and their interactions. Moreover, these methods do not handle missing data or specific data types such as ordinal data.

There are three general multiview learning approaches: weighted view combination, multiview dimension reduction, and subspace learning. Inspired by the promising results of previous multimodal analyses, we aim to explore the effectiveness of three representative methods from the categories, CCA (multiview dimension reduction), MKL (weighted view combination), and CMF (subspace learning), for combining multimodal biomarker features for AD diagnosis. While Zhang et al. [10] and Young et al. [11] used MKL, only linear combination of kernels was explored. CCA and CMF have not been used in the context of fusing biomarkers for AD diagnosis. The fused representation should generalize well to related problems of supervised learning such as classification and unsupervised learning for understanding the association between biomarkers.

Methods

The goals of data fusion are as follows:

1. Reducing the dimensionality of the participating views, so that the fused representation has the most representative components of the individual views.
2. Explaining the nature of relationships between datasets by measuring the relative contribution of each variable to an analysis task.
3. Learning a joint subspace from the different views that supports interpreting the datasets well enough to handle missing data.

We explore the ability of three popular data fusion techniques in attaining these goals. In the implementations, we consider biomarker data as matrices where the rows correspond to subjects and columns to features.

Canonical Correlation Analysis

CCA seeks to find linear projections of two sets of multidimensional variables, so that the projections are maximally correlated [13]. Correlation as a relationship is heavily dependent on the chosen coordinate system; therefore, even if there is a strong linear relationship between two sets of multidimensional variables, the relationship might not be visible as a correlation.

Mathematically, if x and y are two multidimensional random variables with zero mean and $w_a{}^T x$ and $w_b{}^T y$ are their corresponding linear projections, maximizing their correlation, ρ, corresponds to solving Equation (1). If C_{ab} is the cross-covarianve, C_{aa} and C_{bb} are the auto-covariance matrices,

$$\max_{w_a, w_b} \rho = \frac{w_a^T C_{ab} w_b}{\sqrt{w_a^T C_{aa} w_a}\sqrt{w_b^T C_{bb} w_b}} \quad (1)$$

CCA is often formulated as a generalized eigenvalue problem where the maximum correlation corresponds to the largest eigenvalue.

$$\begin{pmatrix} 0 & C_{ab} \\ C_{ba} & 0 \end{pmatrix} \begin{pmatrix} w_a \\ w_b \end{pmatrix} = \rho \begin{pmatrix} C_{aa} & 0 \\ 0 & C_{bb} \end{pmatrix} \begin{pmatrix} w_a \\ w_b \end{pmatrix} \quad (2)$$

Several extensions to the original CCA have been proposed to include more than two views, and to find non-linear relationships between views. Currently, we restrict ourselves to the linear version because it is faster and involves easily interpretable components. The most commonly used approach to include three or more, say p data sources is to sum up the correlations (mCCA). The generalized eigenvalue problem then accounts for maximizing the sum of the correlations. This formulation is depicted in Figure 2.

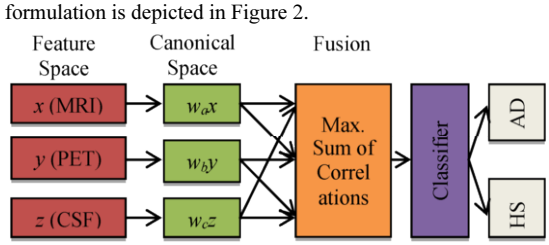


Figure 2. CCA based classification

Tripathi et al. [13] proposed a two step procedure for summing up the correlations. First, the correlations within a data source are removed by a process called whitening. This is done by multiplying the individual data matrices with the square-root of their respective covariance matrices to find components shared between the views. Second, Principal Component Analysis (PCA) is applied to the column-wise concatenation of the whitened data sources. The original data is further projected on to the largest d PCA coefficients. The choice of d is based on the amount of shared variance. The smallest d, after which there is no significant increase in shared variance, is the optimal dimension of of the projection. The projection yields the fused representation which is then used by a classifier to learn the model.

Multiple Kernel Learning

Kernel methods such as SVM, which are based on similarity measures between data points, have been used with great success for dimensionality reduction and classification. Kernelization projects the native space data to a higher dimensional feature space. Non-linear relations between variables in the original space become linear in the transformed space. The projection, ϕ is given by the mapping,

$$\phi: x = (x_1, \dots, x_n) \rightarrow \phi(x) = (\phi_1(x), \dots, \phi_N(x)) \quad (3)$$

To project the data we use the kernel trick, wherein we apply kernel functions, $\kappa_1, \dots, \kappa_p$, to get the corresponding kernel matrices $K_1, \dots, K_p$. Each kernel, $K = <\phi(x), \phi(z)>$ is an inner product of data points. Examples of kernel functions include the linear, radial bias function and others.

Using more than one kernel often produces a better model. In MKL, data is represented as a combination of base kernels [10]. Each base kernel represents a different modality / feature of the entity. MKL seeks to find the optimal combination of the base kernels so that the analysis tasks which follow are benefitted the most. Classification tasks are especially well represented through MKL, as the optimal combination is the one that gives the maximum classification accuracy.

The dual form of MKL optimization, as it is solved by conventional solvers like LIBSVM [14], is

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{m=1}^{p} \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)})$$
$$s.t \quad \sum_{i=1}^{n} \alpha_i y_i = 0; 0 \le \alpha_i \le C; i = 1, \dots, n \quad (4)$$

From a set of n training samples, the features of the i-th sample from the m-th modality are in the vector $x_i^{(m)}$, and its corresponding class label, y_i is either +1 or -1. α's are the Lagrange multipliers which are the variables obtained on converting the primal support vectors to the dual problem. The kernel function applied on each pair of the samples from a modality m, is $\mathbf{k}^{(m)}$. The weights on the m-th modality kernel, represented as β_m are optimized using a grid search or as a separate optimization problem with fixed α. For each new test sample, s, the kernel functions are computed against the training samples. The MKL overview is depicted in Figure 2.

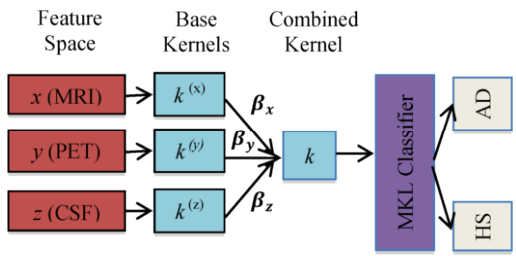


Figure 3. MKL based classification

Zhang et al. [10] and Young et al. [11] used coarse grid search and likelihood maximization approaches respectively, to find the optimal kernel weights, β. They used only one linear base kernel for each of the feature sets and constrain the β's to sum to 1 ($\beta: \|\beta\|_1 = 1; \beta \ge 0$). This however may yield sparse solutions with certain kernels not being well-represented. Recent research has shown that including the base data sets in more than one kernel each differing in their selection of kernel parameters, improves performance [15]. Regularized MKL based on $l2$ norm ($\beta: \|\beta\|_2 = 1; \beta \ge 0$) and $l12$ mixed norm ($\beta: \|\beta\|_2 \le 1; \beta \ge 0$) for constraining β have been proposed. Though $l2$-regularized MKL yields non-sparse solutions, it no longer remains a convex optimization problem and hence is difficult to solve as the sample size increases. $l12$-regularized MKL involves more than one base kernel from a single modality. It enforces sparsity across modalities, while allowing more than one discriminative kernel to be chosen from the same modaility. In other words, there is sparsity across modalities and non-sparsity within modalities, thereby making it a convex optimization problem.

Collective Matrix Factorization

CMF is a technique in relational learning for predicting the unknown values of a relation, given a database of entities and their relations. It learns the low-rank approximations of the matrices which share entities [16].

Given a set of M matrices which describe the relations among E entities, CMF approximates them to low-rank factorizations. The matrices are approximated as a rank-L product and additional row and column bias terms. If r_m and c_m are the entity sets corresponding to the row and column respectively of the m-th matrix, on factorization, its element in the i-th row and j-th column is represented as:

$$x_{ij}^{(m)} = \sum_{l=1}^{L} u_{il}^{r_m} u_{lj}^{c_m} + b_i^{(m,r)} + b_j^{(m,c)} + \varepsilon_{ij}^{(m)} \quad (5)$$

Where, $[u_{ik}^{(e)}]$ is the rank-L approximation of entity set e, $b_i^{(m,r)}$ and $b_j^{(m,c)}$ are the row and column biases respectively and $\varepsilon_{ij}^{(m)}$ is the element-wise noise. The matrices which share the same entity set share the same low-rank matrix approximation. Recent works arrange all the M matrices into a large square grid, whose dimension is the sum of cardinalities of all the entity matrices. However, in the resulting symmetric matrix, Y, only blocks corresponding to the M matrices are observed and the rest of the elements are left unobserved. The CMF model is then formulated as a symmetric matrix factorization,

$$Y = UU^T + \varepsilon \quad (6)$$

where, U is the column-wise concatenation of different $[u_{ik}^{(e)}]$ matrices and bias terms are dropped for simplicity [16].

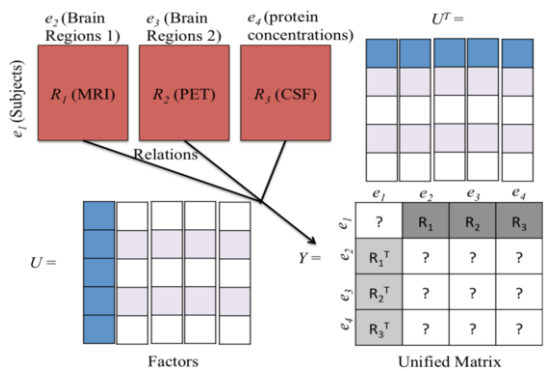


Figure 4. CMF based modeling

Experiments

We implemented the multimodal fusion approaches described above, for integrating the MRI, PET and CSF biomarkers. We used the fused representation to classify a selected study group into patients with AD from healthy subjects (HS). The fused approach is considered successful if the classification task is performed with greater accuracy along with better precision and recall against unimodal classifications. Along with the unimodal approaches, we evaluated the classification of a concatenated data vector comprising data from the three modalities and used it as a baseline study.

The three modalities: MRI, PET and CSF, complement each other in the information they hold [10]; this enables us to draw better insights in a classification task. We used these three biomarkers specifically because, as shown in Figure 1, they compare better than the others in identifying AD early on.

Data

The data for evaluation was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [17]. We worked on baseline MRI images and FDG-PET images that were acquired within 30-60 min post injection. The image details are available at the ADNI website: http://adni.loni.usc.edu/. MMSE (0-30) score of >=27 and CDR (0-3) of 0 are considered normal. The demography of the subjects that we considered are shown in Table 1.

Table 1– Subject Demography

	AD (n = 51; 18F/33M)			HS (n = 52; 18F/34M)		
	Mean	SD	Range	Mean	SD	Range
Age	75.2	7.4	59-88	75.3	5.2	62-85
MMSE	23.8	2.0	20-26	29	1.2	25-30
CDR	0.7	0.3	0.5-1	0	0	0

Preprocessing

The sequence of steps for processing the MRI images included setting the origin to the Anterior Commissure (AC), correcting intensity inhomogeneities, and performing skull stripping. As grey matter atrophy is a prominent feature in AD patients, we segmented the images into grey, white matter and the CSF. This segmentation and the subsequent steps were done using the Statistical Parametric Mapping (SPM) 8 toolbox [18]. To standardize the images of all the subjects, they were normalized to a study specific template created by the SPM DARTEL toolbox [19]. The PET images were co-aligned to the corresponding MRI image using SPM8. Masks of 83 brain regions enlisted in the atlas prepared by Kabani et al. [20] were created using a tool called WFU-PickAtlas [21]. These masks were imposed on the segmented gray matter and PET images to obtain the regional grey matter volume and the average intensity measurements respectively. Thus, we obtained a 1×83 sized feature vector per subject for each of the imaging modalities. The CSF values obtained from ADNI were represented as a 1×3 sized vector per subject representing the total tau, Aβ42 and p-tau values respectively.

We implemented CCA and MKL fusion methods in MATLAB and used the R library 'CMF', for CMF. We tested the individual modalities and the concatenated feature vector (baseline) on the following classifiers:

SVM –This discriminative classifier is accepted to be standard for binary classification. We used the popular LIBSVM [14] tool for our experiments. With unimodal data we used an RBF kernel with default parameters.

GP –We used the GPML toolbox [22] and followed Young et al. [11] for the choice of covariance, mean, likelihood and inference functions.

RF – As an ensemble classifier, we used a MATLAB version of R language's RF library. The number of trees in the classifier were varied according to the dimensionality of the dataset under consideration.

Each method was tested using 10-fold cross validation, categorizing subjects into ten groups based on a random permutation. Nine groups were used for the learning phase and the remaining group formed the test set. The accuracy, precision and recall of the classification tasks were studied. Three prior works in multimodal AD classification were reimplemented and tested with our dataset.

Results

The results of our experiments are tabulated in Table 2. It is evident that a simple concatenation of the feature vectors (SVM (c), GP (c) and RF (c)) provides better classification results than unimodal tests. Prior multimodal biomarker based methods [10, 11, 12] have better classification accuracy than the baseline study (feature concatenation) as expected. However, the results are even better for classification on the fused representation obtained from the statistical methods like CCA and CMF.

Discussion

The concatenated feature vector consistently performs better across the three types of classifiers than individual biomarkers because of their complementary information. The poor performance of the baseline study in which there is no kernel combination, against the prior multimodal analyses of Zhang et al. and Gray et al. [10, 12], is due to the inclusion of all features and not just those which contribute to classification.

The MKL formulation based on *l12* mixed norm performs worse than Zhang et al.'s [10] but better than Young et al.'s [11] both of which are *l1* norm based. As the mixed norm enforces group sparsity, it chooses features common across all participating modalities. In comparison, *l1* norm and Gray et al.'s [12] RF based method choose features individually across modalities and ignores intermodal relationships. From this we understand that the common feature constraint may overlook certain modality specific features aiding classification.

mCCA and CMF both perform better than the rest of the methods. These methods learn a generic model of the biomarkers, not specific for classification. But they perform the best on classification task as well. This is because the generic model learnt from these techniques is built only on those relevant features or components that are statistically dependent across modalities. Though mCCA in its current form is incapable of handling missing entries it may be extended to handle them. CMF performs slightly poorer than mCCA in the classification task but is the most generic model.

The three methods compare as follows, with respect to achieving the goals of data fusion:

1. mCCA is effective in data exploration to find if there are any associations between the data sources. It saves what is shared between the views and ignores variations within, thereby achieving goals 1 and 2.
2. If the goal is only supervised learning, MKL methods, *l12* and *l1* based optimization [10, 11], can be applied directly as they learn the most distinguishing multimodal features, satisfying goal 1. However, such methods fail when there is missing data. Moreover, these methods lack a proper generative model for each view, and hence cannot be used for the task of understanding the data.
3. CMF handles missing entries by treating them as test data and allows multiple likelihood functions for modeling the data. The benefit of using CMF is that it identifies common factors shared between matrices and factors specific to individual matrices. Matrix factorization results in dimensionality reduction and thus satisfies the three goals of data fusion.

Conclusion

We examined multimodal data fusion on a dataset consisting of heterogeneous biomarker data. We used three categories of

fusion methods based on CCA, MKL and CMF. Further, we used the resultant fused representation for classifying AD patients. We found that classifying based on the fused representation that preserves intermodal relationships yields better results than unimodal classification. Amongst the three methods, mCCA gives the best accuracy on our dataset closely followed by the CMF based method.

Table 2– Region of Interest Based Classification

			Precision		Recall	
Data	**Method**	**Acc.**	**AD**	**HS**	**AD**	**HS**
MRI	SVM	82.7	86.7	79.3	82.8	78.4
	GP	81.5	86.8	78.4	76.8	84.6
	RF	82.7	86.5	85.4	81.7	81.6
PET	SVM	85.5	88.4	86.4	85.3	84.3
	GP	82.6	84.2	81.3	82.1	83.1
	RF	81.5	81.4	87	94	73.6
CSF	SVM	80.6	81.2	81.3	83.1	81.6
	GP	81.5	83.2	83.9	85.9	78.9
	RF	81.6	83.7	81.9	82.6	82.9
MRI +	[10]	92.4	87.9	86.4	88.1	84.7
PET +	[11]	87.5	87.9	89.6	87.7	84.6
CSF	[12]	91.5	91.7	91.7	93.2	90.6
	SVM (c)	86.5	88.6	90	88.2	80.4
	GP (c)	89.3	89.6	93.7	91.5	82.9
	RF (c)	90.5	88.3	96.6	95.5	83.6
	mCCA	**95.1**	**94.8**	**97.1**	**96**	**94.2**
	l12-MKL	88.4	86.6	92.2	92.9	83.9
	CMF	94.4	84.5	96.3	87.3	87.3

Acknowledgments

We would like to thank Dr. Kandiah Nagaendran, Dr. Ming-Ching Wen, and Dr. Tchoyoson Lim, and their teams at the National Neuroscience Institute, Singapore, for providing the medical background and guidance. This project was supported by an Academic Research Grant No. T1 251RES1211 and a research scholarship from the Ministry of Education, Singapore. Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu/).

References

[1] Atrey P, Hossain M, El Saddik A, and Kankanhalli M. Multimodal fusion for multimedia analysis: a survey. Multimedia Systems, 2010; 16(6): 345–379.

[2] Alzheimer's Disease International. The Global Impact of Dementia 2013–2050, 2013. Available from: https://www.alz.co.uk/research/GlobalImpactDementia2013.pdf (accessed 11 December 2014).

[3] National Institute of Aging. Advances in Detecting Alzheimer's Disease. 2011-2012 Alzheimer's Disease Progress Report, 2012. Available from: http://www.nia.nih.gov/alzheimers/publication/2011-2012-alzheimers-disease-progress-report/advances-detecting-alzheimers (accessed 11 December 2014).

[4] Lanckriet GRG, M Deng, N Cristianini, MI Jordan, and WS Noble. Kernel-Based Data Fusion And Its Application To Protein Function Prediction In Yeast. Pacific Symposium on Biocomputing, 2004.

[5] Lee G, Doyle S, Monaco J, Madabhushi A, Fledman M, Master S, and Tomaszewski J. A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology. IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009.

[6] Baez PB, Viadero C, Perez del Pino M, Prochazka A, and Suarez-Araujo C. Humann-based systems for differential diagnosis of dementia using neuropsychological tests. 14th Intl. Conference on Intelligent Engineering Systems, 2010.

[7] Hua XS, and Zhang HJ. An attention-based decisionfusion scheme for multimedia information retrieval. 5th Pacific-Rim Conference on Multimedia, 2004.

[8] Radova V, and Psutka J. An approach to speaker identification using multiple classifiers, IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997.

[9] Wu Y, Chang EY, and Tseng BL. Multimodal metadata fusion using causal strength. 13th annual ACM international conference on Multimedia, 2005.

[10] Zhang D, Wang Y, Zhou L, Yuan H, and Shen D. Multimodal classification of alzheimer's disease and mild cognitive impairment. NeuroImage, 2011; 55(3): 856 – 867.

[11] Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, and Ourselin S. Accurate multimodal probabilistic prediction of conversion to alzheimer's disease in patients with mild cognitive impairment. NeuroImage, 2013; 2(0):735 – 745.

[12] Gray K, Aljabar P, Heckemann R, Hammers A, and Rueckert D. Random forest-based manifold learning for classification of imaging data in dementia. In: Machine Learning in Medical Imaging. Lecture Notes in Computer Science, 2011; 7009: 159–166.

[13]Tripathi A, Klami A, Kaski S. Simple integrative preprocessing preserves what is shared in data sources. BMC Bioinformatics; 2008; 9(1):111.

[14] Chang CC and Lin CJ, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2011; 2(27): 1-27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[15]Xu Z, Jin R, Yang H, King I, Lyu M. Simple and Efficient Multiple Kernel Learning by Group Lasso. International Conference on Machine Learning, 2010.

[16]Klami A, Bouchard G, and Tripathi A. Group Sparse Embeddings in Collective Matrix Factorization. International Conference on Learning Representations, 2014.

[17] ADNI | Alzheimer's Disease Neuroimaging Initiative. http://adni.loni.usc.edu/ (accessed on 11 December 2014].

[18] Friston K, Ashburner J , Kiebel S, Nichols T, and Penny W. Statistical parametric mapping: The analysis of functional brain images. Elsevier, London, 2006.

[19] Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage, 2007; 38(1): 95 – 113.

[20] Kabani N, MacDonald D, Holmes CJ, and Evans A. A 3D atlas of the human brain. Neuroimage 7, 1998; S717.

[21] Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. NeuroImage, 2003; 19: 1233– 1239.

[22] Rasmussen CE and Nickisch H. Gaussian Processes for Machine Learning (GPML) Toolbox. J. Mach. Learn. Res. 11, 2010; 3011-3015.

Address for correspondence

Parvathy Sudhir Pillai (parvathy@comp.nus.edu.sg)

Medical Computing Laboratory, School of Computing, National University of Singapore, Computing 1, 13 Computing Drive, Singapore-117417

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-736

Towards Constructing a New Taxonomy for Psychiatry Using Self-reported Symptoms

Jessica Ross [a], Thomas Neylan[b], Michael Weiner[b,c], Linda Chao[c], Kristin Samuelson[d,e], Ida Sim[f]

[a] *Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA*
[b] *Department of Psychiatry, University of California, San Francisco and San Francisco VA Medical Center, San Francisco, CA, USA*
[c] *Department of Radiology, University of California, San Francisco and San Francisco VA Medical Center, San Francisco, USA*
[d] *San Francisco VA Medical Center, San Francisco, CA, USA* [e] *Alliant International University, San Francisco, CA, USA*
[f] *Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA*

Abstract

The Diagnostic and Statistical Manual (DSM) has served as the gold standard for psychiatric diagnosis for the past several decades in the USA, and DSM diagnoses mirror mental health and substance abuse diagnoses in ICD-9 and ICD-10. However, DSM diagnoses have severe limitations when used as phenotypes for studies of the pathophysiology underlying mental disorders, as well as for clinical treatment and research. In this paper, we use a novel approach of deconstructing DSM diagnostic criteria, and using expert knowledge to inform feature selection for unsupervised machine learning. We are able to identify clusters of symptoms that stratify subjects with the same DSM disorders into cohorts with increased clinical and biological homogeneity. These findings suggest that itemized self-report symptom data should inform a new taxonomy for psychiatry, and will enhance the bi-directional translation of knowledge from the bench to the clinic through a common terminology.

Keywords:

Psychiatry; Taxonomy; Unsupervised Learning; Clinical and Translational Informatics

Introduction

The Diagnostic and Statistical Manual of Mental Illness (DSM-V) is the current clinical psychiatric classification system in the United States for mental illness and substance abuse disorders [1]. When a patient is given a DSM diagnosis, the clinician asks the patient about the presence or absence of numerous symptoms, and then uses multiple DSM algorithms to determine the best diagnosis for the patient. There is a growing acceptance that DSM syndromes have severe limitations when used as phenotypes for identifying biomarkers associated with mental illness. Whether the symptom data that is ascertained during the diagnostic process can be used more effectively as phenotypes for biological inquiry into the pathophysiology of mental illness remains unclear. The purpose of this study is to determine if atomic level symptom data that is regularly ascertained by mental health practitioners can be effectively used to develop phenotypes that can identify clinical cohorts of psychiatric patients with pathophysiological homogeneity. As these data are already ascertained as part of the normal clinical workflow in psychiatry, and are frequently reported in the medical record, they can theoretically be used to phenotype clinical populations, facilitating cohort identification for biomarker identification, and enabling observational studies on outcomes effectiveness in different cohorts. These studies might in turn lead to the identification of biomarkers that can be used to stratify psychiatric populations in the future, improving diagnosis, treatment, and outcomes in these populations.

The current psychiatric classification system, the DSM, was initially developed after World War II, motivated to a large degree by the need for the Armed Forces to monitor the prevalence of and treat mental illness in soldiers who returned from the War. DSM I in 1952, and DSM II in 1968, were both rooted in psychodynamic psychiatry, in line with the legacy of psychoanalysis (e.g. beginning with Freudian Theory). With DSM III, the decision was made to take an "atheoretical" approach, and develop a classification system based on patterns of symptoms that clustered together, as agreed to by expert consensus in committee meetings, without the use of empirical data, and divorced from explanatory models of the etiologies of these disorders [2]. Studies showed that this new approach vastly improved inter-rater reliability with respect to the psychiatric diagnosis patients were given [3]. While the approach presumed that the categories reflected underlying pathological processes, no quantitative data or biological markers have been included in the DSM thus far.

Since 1994, the DSM-IV has served as the gold standard in psychiatric treatment and research [4]. Though DSM-5 was officially released in May of 2013, the vast majority of USA clinicians are still transitioning from DSM-IV to DSM-5. As such, DSM-IV disorders have served as phenotypes by which clinicians diagnose and treat patients and obtain reimbursement, as well as phenotypes for basic science, translational, and clinical research carried out in the domain of psychiatric illness. Furthermore, DSM-IV syndromes have been shown to have a great deal of similarity to ICD-9 and ICD-10 diagnoses, so the results of studies using the DSM are highly generalizable to populations worldwide [5].

Much has been written about the limitations of the DSM-IV with respect to the use of its syndromes as phenotypes for use in both clinical practice and research. Below, we will briefly review the most salient points as they relate to phenotypic heterogeneity, limitations of syndromes in identifying biomarkers, and informatics challenges faced when attempting to construct phenotypes that define more clinically and biologically homogenous groups from symptom data ascertained using DSM-IV syndromes as phenotypes.

DSM-IV defines 137 Syndromes/Disorders across 15 categories. Syndromes are by definition binary, and there are no signs (objective or quantitative markers) associated with DSM-IV syndromes. Symptom overlap (e.g. sleep disturbances, guilt) across syndromes (e.g. Major Depressive Disorder (MDD), Post Traumatic Stress Disorder (PTSD)), resulting in clinical presentations where patients are defined as having several "comorbid" psychiatric disorders, complicating

diagnosis and treatment for the clinician, with a paucity of evidence based treatment algorithms for multiple comorbid psychiatric disorders.

DSM-IV phenotypes also have limited clinical utility, as two patients with the same DSM-IV Syndrome may not share any of the same symptoms. Due to the structure of DSM-IV Syndromes, many different combinations of symptoms meet the criteria for the same Syndrome. For example, with Major Depressive Disorder (MDD), "depressed mood" or "anhedonia" AND 5 out of 9 other symptoms are required, leading to 112 different possible symptom presentations. As some of the symptoms are underspecified (e.g. sleep disturbance may refer to hypersomnia or insomnia), it is possible for two patients to have MDD without sharing ANY of the same symptoms. Of note, this lack of syndromic specificity remains present in DSM-5, as many major syndromes, including MDD have not been altered at all for the new version.

This phenotypic heterogeneity also has severe limitations for clinical and translational research. Clinical trials have shown that several classes of medications that are equally efficacious across several syndromes, and are used clinically across several DSMIV higher level classes(e.g. Mood Disorder and Anxiety Disorders, Mood Disorders and Psychotic Disorders) arguing for the fact that the overlap of symptoms across these disorders actually reflects the presence of common pathophysiology present across different DSM-IV syndromes.

There has been an abundance of articles discussing how the lack of biological validity underlying the creation of DSM-IV Syndromes has resulted in a lack of robust biomarker findings within syndromes [6]. In addition to the fact that DSM-IV Syndromes have overlapping symptoms, recent studies have confirmed that specific genetic variants are present in subjects with multiple DSM-IV Syndromes, and not in unaffected controls. One recent seminal study showed that variants in the CACNA1C calcium channel are associated with several psychiatric syndromes, including autism spectrum disorder, attention deficit hyperactivity disorder, bipolar affective disorder, and major depressive disorder [7]. Additionally, several other biological markers, for example non-suppression of cortisol by dexamethasone, have been found to be significantly different in affected individuals as opposed to controls, in multiple psychiatric syndromes [8]. However, to our knowledge, there has not been a comparison of psychiatric symptom presentation similarities across these syndromes, with a focus on how this relates to these biomarker differences.

Finally, the current terminology of the DSM-IV nosology makes it difficult to investigate symptoms across syndromes, due to a decision made with DSM-III that no symptom could be replicated in two Syndromes [9]. Though it is unclear if this rule has explicitly followed during the creation of DSM-IV and 5, manifestations of this limitation undoubtedly exist in DSM-IV and DSM- 5, making it challenging at this time to use existing NLP and text-mining methods to automatically identify symptoms across syndromes within narrative text, or even structured interview data. The APA has increasingly published measures to investigate more primary psychiatric symptoms in clinical populations, but the use of these measures has not yet translated into clinical care, nor is it yet reflected in the literature [1].

In view of a preponderance of shortcomings, major stakeholders in mental health treatment and research have publically presented alternate systems for codifying current and future knowledge to inform the understanding of mental illness, most notably Precision Medicine and the Research Domain Criteria (RDoC)) [10]. However, these ideas are still very much in the development stage, and there has been a paucity of discussion as to how these taxonomies will be integrated with clinical psychiatry. Generations of psychiatrists, including those who are currently teaching residents and medical students, have been trained using the DSM, and routinely evaluate patients using symptom level data with the ultimate goal of identifying the most appropriate DSM disorder or disorders for the patient based on the constellation of symptoms present in the clinical presentation.

It is with this background that we propose to identify groups of patients that cluster together based on itemized psychiatric symptom level presentation, regardless of how these symptoms have been used to classify DSM Syndromes up until this point. We then propose to identify biomarkers that are associated with these different clinical symptom presentations to show that self-report symptoms have utility in developing a clinically relevant and biologically valid taxonomy for psychiatry. Few studies have looked at the correlation of psychiatric symptoms with biomarkers, independent of DSM-IV syndromes, across a patient population. The vast majority of studies which have looked at itemized symptom level data in the DSM-IV in any manner, have looked only at a subset of symptoms involved in one to three closely related DSM disorders, most notably within childhood disorders such as autism and ADHD, and cognitive disorders including Alzheimer's. There are two studies we know of that have looked at the DSMIV symptoms in total, but with different foci than described in this proposal [11, 12] We have not been able to identify any studies that have looked in total at all symptoms across all principal DSM Syndromes within a dataset where patients carry the diagnosis of more than one DSM Syndrome.

Studies have shown that it is possible to classify psychiatric patients into groups of subjects with increased pathophysiological and disease course homogeneity using machine learning algorithms with various types of biomarker data [13]. This classification is possible even without full understanding of the pathophysiology underlying the disorder. These methods also suggest a direct translation from academic research to clinical practice, where they may ultimately be used to diagnose patients in the clinical setting.

We present the analyses of a multi-modal dataset comprised of combat veterans, 33 % of whom have been given a DSMIV diagnosis of PTSD. We used unsupervised learning methods, including hierarchical cluster analysis and K means clustering, with the goal of identifying robust groups of patients that cluster together with respect to clinical symptoms and biomarkers.

Materials and Methods

Data

We conducted secondary analyses ascertained through VA and DoD funding at the San Francisco Veterans Administration Health Center. Details of the Gulf War (GW) study have also been published [14, 15]. In brief, there were 292 Gulf War Veterans, 85% of whom were male, and 75% who were combat trauma exposed. Given the small sample size and the relatively large number of clinical and biological markers to be used in the analyses, only the males were included in this study. 50% of those included had comorbid DSM-IV diagnoses of Lifetime Alcohol Abuse or Dependence, and 40% had comorbid DSM-IV diagnoses of Major Depressive Disorder.

The four main data types used in this study are self-report clinical symptom data, and three different modalities of biomarker data: neuropsychological data, magnetic resonance imaging data, and neuroendocrine data. Ascertainment of these data are briefly described below.

The self report clinical data that was used consists of 136 clinical self-reported clinical features, 21 from the Beck Depression Inventory (BDI;, a scale developed and used to diagnose Major Depressive Disorder, 19 from the Clinician's Assessment for PTSD (CAPS), a scale used to diagnose PTSD in clinical populations and 90 items from the Symptom Check List -90 (SCL-90), a clinical questionnaire developed to measure psychiatric symptoms across 10 psychiatric domains, including mood, psychosis, anxiety, eating, and sleep disorders. Also included were 5 binary measures of significant substance use or dependence, for alcohol, cannabinoids, stimulants (e.g. methamphetamine), cocaine, and hallucinogens.

The ascertainment of neuropsychological and imaging data was described in detail in prior publications. Briefly, the neuropsychological battery used assessed three domains of cognitive functioning: verbal memory; visual memory and visual–spatial skills; and attention, working memory, and processing speed with 10 normalized scores reported for each subject. Subjects were studied with MRI; multislice 1H MRSI measures were ascertained. Hippocampal determination was based on MPRAGE images and carried out semi-automatically using a high dimensional brain-mapping tool and values were divided by total intracranial volume for normalization. Subjects also completed a low-dose (0.5 mg) dexamethasone (DST) suppression test challenge using salivary cortisol [8] , and area under the curve (AUC) for both day 1 and day 2 of the test were used in these analyses. Baseline serum cortisol level was also ascertained.

Feature Selection

A hybrid approach was attempted to reduce redundancy across clinical features: Features were first mapped to each other if a clinical expert (JR) determined the concepts they were identifying as being clinically equivalent (e.g. expected to result in the same answer if both questions were asked to the same patient in a clinical interview). A correlations matrix was also created for all clinical features, to identify features with a correlation coefficient of $\geq$ .90. Interestingly, there was no intersection between the two methods, and the decision was made to remove a feature based on correlation coefficient $\geq$ .90, resulting in 124 total clinical features that were ultimately used for the subsequent analyses.

Ideally we would have been able to use symptom data from the Structured Interview for DSM (SCID), the tool that was used to initially identify DSM Diagnoses. However, the data in this study has been generated without the SCID data, as we only currently have composite level syndrome data available from the SCID. We therefore mapped the complete CAPS, BDI, SCL-90, and Alcoholic and substance abuse portions of the SCID to all the SCID questions, and estimate that our data cover approximately 85 % of all symptoms ascertained in a formal SCID, with predominantly 75% of psychotic symptom data and 50% of OCD symptom data unaccounted for. However, as individuals were excluded from our study with a history of psychotic symptoms, we estimate that missing relevant symptomatic data are minimal.

Statistical Analyses

Individuals with > 15% data missing in any measures (clinical, imaging, endocrine, neuropsychiatric), were not included for any analyses that required that data type. Missing values were filled with median imputation. All data used in the analyses was standardized and normalized.

All statistical analyses were performed using R statistical programming software [16]. We used hierarchical clustering (*hclust*) with Ward's distance measurement. As multiscale bootstrap resampling *(pvclust)* provided no significant p-values (e.g. p<. 05) for any individual clusters, clusters were delineated using a dendrogram to visualize vertical distances in branching across clusters. We then performed K-means nearest (*kmeans*) neighbor clustering three different times, as clusters delineated with the *kmeans* algorithm are nondeterministic. We identified the optimal number of clusters from the K-means analyses by finding the bend in the plot of the within group sum of squares versus number of clusters. We then compared cluster solutions across all three K-means results with the hierarchical clustering results to determine robustness of clusters. Finally, we performed descriptive statistical analysis on imaging, neuropsychiatric, and neuroendocrine measures using ANCOVA to determine differences in biomarkers across clusters.

Results

Hierarchical agglomerative clustering using 124 features using Ward's method for determining distance between individuals produced the results shown in Figure 1. Visual inspection of dendrogram branching identified two, three, or five natural clusters. Multiscale bootstrap resampling provided no significant p-values (e.g. p<. 05) for any individual clusters. Three independent runs of K-nearest neighbor analysis also revealed that data ideally clustered into 5 groups that had a good to great degree of similarity with the hierarchical clusters (Cohen's Kappa Values .904, .897, .702).

Given the similarity of the multiple solutions and the determinism in using agglomerative hierarchical clustering, the five clusters delineated through hierarchical clustering were chosen for further analyses of biomarkers across clusters. Ancova with adjustment for age was calculated for all biomarkers in the data set. The most significant p-values, values without adjustment for multiple testing are shown in Table 1. In an effort to account for variance, a linear model was constructed using all variables that had unadjusted p-values approaching a level of significance in a Mancova analysis, with results for differences in biomarkers across clusters shown in Table 2.

To further gain insight into differences between the five clusters, we looked at the absolute values of individual symptoms across the five clusters, as shown in Figures 2a, 2b, and 2c. We also graphed mean values of all biomarkers across the five clusters to facilitate description of these five clusters both from the perspective of biomarkers, and of clinical symptoms. Graphs for the six biomarker variables that showed significance in the Mancova analyses, are shown in Figures 3a through 3f.

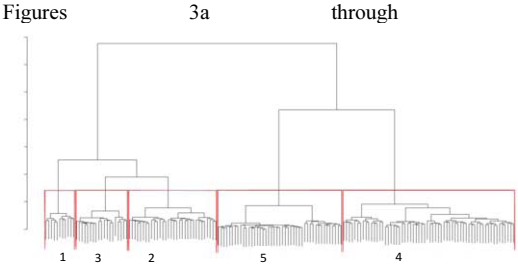


Figure 1-Hierarchical Clustering of Itemized Symptom Data

Table 1-Ancova of Clusters with Significant Biomarkers

Dependent Variable	Ancova p-value
Right Hippocampal Volume	.057
Dexamethasone Suppression Test	.004
Executive Functioning	.057
Processing Speed	.030
Manual Dexterity	.007
Performance IQ	.016

Table 2-Mancova of Clusters using Significant Biomarkers

Biomarker	Mancova p-value
DST:Area Under the Curve day 1	.910
DST:Area Under the Curve day 2	.008
Baseline Serum Cortisol	.222
Memory	.024
Executive Functioning	.003
Learning	.039
Attention	.034
Processing Speed	.003
Manual Dexterity	.056
Ataxia	.004
Verbal IQ	.003
Performance IQ	.000
Right Hippocampal Volume	.417
Left Hippocampal Volume	.517

Discussion

In this study we show that statistical learning methods can be used with atomic level psychiatric symptom data to identify groups of patients with phenotypic and biological homogeneity. Perhaps the most surprising finding is that the clusters delineated are not associated with any "type" or "types" of symptoms as would have been expected from the DSM classification system. Instead, the 5 clusters are defined by the intensity of the majority of symptoms across all questionnaires, and hence, DSM Syndromes. In fact, with the exception of two neuro-vegetative symptoms associated with depression and one paranoia symptom, the subjects in Cluster 1 on average have worse values for all of the psychiatric symptoms than subjects in all of the other clusters. As would be expected, these subjects have a higher mean number of DSM–IV diagnoses, yet no single or even group of co-morbid diagnoses can partition them in the intuitive manner that hierarchical clustering is able to.

Cluster analyses delineated 3 clusters of patients that each contain significant numbers of individuals who meet criteria for PTSD (clusters 1, 2, and 4), as seen in Table 3. However, individuals in cluster 4 have much lower levels of subjective psychiatric symptoms than individuals in clusters 1 or 2, and significantly different (and arguably more "normal") biological markers across the neuropsychiatric data, imaging data, and neuroendocrine data. These findings are in line with the growing consensus that DSM phenotypes are underspecified, and lump together groups of patients that are too heterogeneous to use in clinical, translational, or basic science studies. While this study is obviously limited by the very small sample size and exclusion of female subjects, the fact that biologically and clinically relevant clusters were identified, provides support for the hypothesis that using data from larger psychiatric cohorts with these methods will greatly increase our current knowledge base in the psychiatric domain.

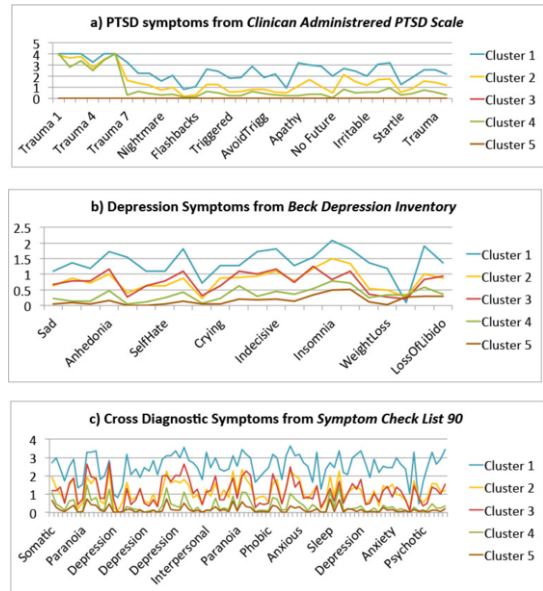


Figure 2- Mean itemized symptom data for each cluster

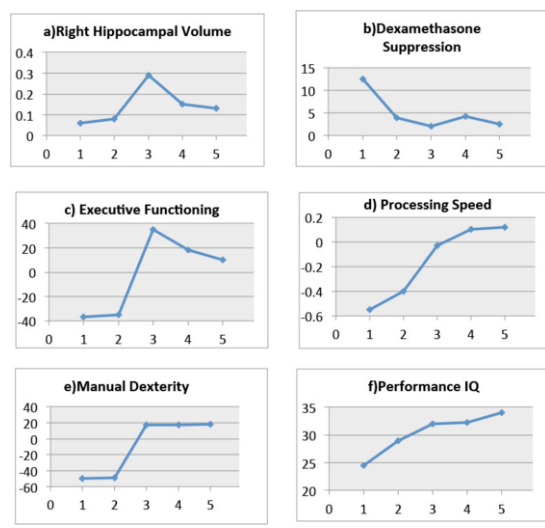


Figure 3- Normalized Mean Biomarker Values By Cluster: a) Right Hippocampal Volume, b)DST Area Under the Curve,c-e)Standardized Executive Functioning, Processing Speed, and Manual Dexterity, f)Performance IQ

The use of DSM diagnoses does not correlate with these clusters, but symptom data analogous to that used to delineate DSM diagnoses were used in these analyses. Therefore, these findings propose a method by which clinical stakeholders and basic science researchers in the mental health domain can communicate to delineate a taxonomy for psychiatry that is ultimately both clinically and biologically useful. Arguably, phenotype definition is the most pressing issue in psychiatry

Table 3-DSM-IV Diagnoses by Cluster Number

Cluster	Tot. #	PTSD	MDD	ALC	PTSD MDD	PTSD ALC
1	11	11	5	4	5	2
2	32	12	4	15	0	1
3	19	0	6	4	0	0
4	61	16	2	14	1	0
5	45	1	1	4	0	0

to date. DSM-5, released in May 2013, remains the clinical bible for psychiatrists, and its diagnoses, which map to ICD-9 and ICD-10, are necessary for billing insurance companies and Medicare. DSM-5 therefore remains the standard in medicine, and continues to be taught to medical students and residents. Yet one month prior to the release of DSM-5, the NIMH proclaimed that studies using DSM phenotypes will no longer be a priority for funding, strongly encouraging the use of the RDoCs [10]. While a biologically based taxonomy is the ultimate goal for psychiatry, as with all other domains of medicine, clinical psychiatry is far from being grounded by mechanistic biomarkers. The findings in this paper that individual symptoms that are regularly ascertained clinically are associated with changes across multi-modal biomarkers, gives support to the notion that symptoms may still be useful in psychiatry to both clinicians and researchers. In fact, there may be data within the multitude of large datasets already ascertained, that can be used to develop symptom based psychiatric phenotypes that can identify clinically and biologically homogeneous subpopulations, both for diagnostic purposes, and that will have clinical utility and be able to further inform phenotype construction in a biologically based taxonomy.

It seems likely in the near future that patients will continue to present with self-report symptoms that will guide the evaluation and treatment of patients. While currently clinicians work towards identifying the most appropriate DSM diagnosis or diagnoses for each client to guide treatment, it may require few extra resources to obtain more extensive detailed symptom level data, that can be used with statistical learning methods to cluster clients into groups with increased biological and clinical homogeneity. These cohorts can be followed in observational studies in outcomes effectiveness, and potentially be recruited to participate in biomarker studies. Additionally, in the age of the Internet, ascertaining self-report symptoms require few resources and can be done remotely.

In conclusion, this study provides novel findings indicating that self-report psychiatric symptoms may be used to identify clinically and biologically homogeneous psychiatric populations. These symptom data can be used when constructing a biologically valid taxonomy for psychiatry and will enhance the bi-directional translation of knowledge from the bench to the clinic through a common terminology. Continued work into feature selection and interpretation of results with these data will facilitate the construction of a psychiatric taxonomy that is biologically valid and translates to clinical psychiatric care.

Acknowledgments

Jessica Ross was supported by the Veteran's Administration Advanced Fellowship in Psychiatry/Neurosciences.

References

[1] American Psychiatric Association., American Psychiatric Association. DSM-5 Task Force. Diagnostic and statistical manual of mental disorders : DSM-5. 5th ed. Arlington, VA: American Psychiatric Association; 2013.xliv, 947p. p.

[2] American Psychiatric Association. Task Force on Nomenclature and Statistics., American Psychiatric Association. Committee on Nomenclature and Statistics. Diagnostic and statistical manual of mental disorders. 3d ed. Washington, D.C.: American Psychiatric Association; 1980. 494 p. p.

[3] Segal DL, Hersen M, Van Hasselt VB. Reliability of the Structured Clinical Interview for DSM-III-R: an evaluative review. Comprehensive psychiatry. 1994;35(4):316-27.

[4] American Psychiatric Association., American Psychiatric Association. Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders : DSM-IV. 4th ed. Washington, DC: American Psychiatric Association; 1994. xxvii, 886 p. p.

[5] Jablensky A. The nature of psychiatric classification: issues beyond ICD-10 and DSM-IV. The Australian and New Zealand journal of psychiatry. 1999;33(2):137-44.

[6] Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. Br J Psychiatry. 2010;196(2):92-5.

[7] Cross-Disorder Group of the Psychiatric Genomics C, Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381(9875):1371-9.

[8] Kirschbaum C, Hellhammer DH. Salivary cortisol in psychobiological research: an overview. Neuropsychobiology. 1989;22(3):150-69.Weed L.

[9] Franklin CL, Zimmerman M. PTS and major depressive disorder: investigating the role of overlapping symptoms in diagnostic comorbidity.The Journal of nervous and mental disease.2001;189(8):548-51.

[10]Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. The American journal of psychiatry. 2014;171(4):395-7.

[11]Wessman J, Paunio T, Tuulio-Henriksson A, Koivisto M, Partonen T, Suvisaari J, et al. Mixture model clustering of phenotype features reveals evidence for association of DTNBP1 to a specific subtype of schizophrenia. Biol Psychiatry. 2009;66(11):990-6.

[12]Borsboom D, Cramer AO, Schmittmann VD, Epskamp S, Waldorp LJ. The small world of psychopathology. PloS one. 2011;6(11):e27407.

[13]Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch Gen Psychiatry. 2009;66(7):700-12.

[14]Apfel BA, Ross J, Hlavin J, Meyerhoff DJ, Metzler TJ, Marmar CR, et al. Hippocampal volume differences in gulf war veterans with current versus lifetime posttraumatic stress disorder symptoms. Biol Psychiatry.69(6):541-8.

[15]Samuelson KW, Neylan TC, Metzler TJ, Lenoci M, Rothlind J, Henn-Haase C, et al. Neuropsychological functioning in posttraumatic stress disorder and alcohol abuse. Neuropsychology. 2006;20(6):716-26.

[16]Team RDC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria 2011.

Address for correspondence

J.A.Ross. Assistant Clinical Professor, Department of Psychiatry, University of California, San Francisco 1001 Potrero Avenue, 7M, San Francisco, CA, 94110. Email: Jessica.Ross@ucsf.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-741

A Multi-Relational Model for Depression Relapse in Patients with Bipolar Disorder

Rogerio Salvini[a,c], Rodrigo da Silva Dias[b], Beny Lafer[b], Inês Dutra[c,d]

[a] *Instituto de Informática, Universidade Federal de Goiás, Goiânia, GO, Brazil*
[b] *Bipolar Research Program, Department and Institute of Psychiatry, University of São Paulo Medical School, São Paulo, Brazil*
[c] *CRACS & INESC TEC-Porto LA, University of Porto, Porto, Portugal*
[d] *University of Wisconsin-Madison, Madison, WI, USA*

Abstract

Bipolar Disorder (BD) is a chronic and disabling disease that usually appears around 20 to 30 years old. Patients who suffer with BD may struggle for years to achieve a correct diagnosis, and only 50% of them generally receive adequate treatment. In this work we apply a machine learning technique called Inductive Logic Programming (ILP) in order to model relapse and no-relapse patients in a first attempt in this area to improve diagnosis and optimize psychiatrists' time spent with patients. We use ILP because it is well suited for our multi-relational dataset and because a human can easily interpret the logical rules produced. Our classifiers can predict relapse cases with 92% Recall and no-relapse cases with 73% Recall. The rules and variable theories generated by ILP reproduce some findings from the scientific literature. The generated multi-relational models can be directly interpreted by clinicians and researchers, and also open space to research biological mechanisms and interventions.

Keywords:

Bipolar Disorder; Depression Relapse; Multi-relational Model; Inductive Logic Programming; Machine Learning.

Introduction

The Diagnostic and Statistical Manual of Mental Disorders (DSM), published by the American Psychiatric Association, organizes all psychiatric disorders and other problems into five different categories, or axes. Axis I disorders, the focus of this work, includes all psychological diagnostic categories such as depression, anxiety disorders, and bipolar disorder, among others, except mental retardation and personality disorders that fall into Axis II. A psychiatric disorder involving both manic and depressive episodes, Bipolar Disorder (BD), which falls into the Axis I category, is a chronic and disabling disease that usually appears around 20 to 30 years old. Patients who suffer with BD may endeavor for years to achieve a correct diagnosis, and only 50% of them generally receive adequate treatment. Commonly, the mood episodes lead to personal and professional problems, and due to the disease's progression character, the euthymic periods (without mood episodes) become gradually shorter [1]. A WHO study [2] evidenced the devastating impact of BD: all around the world, approximately 29.5 million people present BD, which occupies the 12th position relative to the major causes of moderate to severe incapacity. With 22.2 million people in such conditions, this study also reveals that 2.5% of the years of incapacity are related to BD episodes. Relapse prevention is the main target in BD treatment, since the relapse rate is around 50% at one year and 70% at four years treatment [1, 3, 4].

Due to this alarming picture, a number of researchers started to explore the aspects underlying the BD outcome, in particular, genetic, neurobiological, environmental and phenomenological factors, as well as the interactions between them. As BD presents multi-factorial and multidimensional characteristics, there appears to be an urgent necessity of new techniques to investigate patterns associated with the course of this disease, especially the computational and mathematical ones. Machine learning and data mining bring the possibility of new hypotheses discovery through the exploration of datasets (new or not), neuroimaging and biomarkers, thus opening space to the research of new biological mechanisms and interventions [5, 6].

In fact, machine learning have been already applied to the study of BD. Mourão-Miranda *et al*. [7] have applied Gaussian Process Classifiers (GPC), a machine learning approach that assigns a predictive probability of group membership to an individual person, to differentiate groups and to identify those at-risk adolescents most likely to develop future Axis I disorders. The work was done collecting information from functional magnetic resonance imaging after the teens performed two emotional face gender-labeling tasks (happy/neutral; fearful/neutral). Using GPC, neural activity to neutral faces presented during the happy/neutral experiment accurately and significantly differentiated groups, achieving 75% accuracy (sensitivity=75%, specificity=75%). More recently, Schnack *et al*. [8] used Support Vector Machines (SVM) to distinguish among three groups of patients. They scanned 66 schizophrenia patients, 66 patients with bipolar disorder and 66 healthy subjects on a 1.5 Tesla magnetic resonance imaging scanner. Three SVM were trained to separate patients with schizophrenia from healthy subjects, patients with schizophrenia from those with bipolar disorder, and patients with bipolar disorder from healthy subjects, respectively, based on their gray matter density images. The predictive power of the models was tested using cross-validation and in an independent validation set of 46 schizophrenia patients, 47 patients with bipolar disorder and 43 healthy subjects scanned on a 3 Tesla magnetic resonance imaging scanner. Schizophrenia patients could be separated from healthy subjects with an average accuracy of 90%. Additionally, schizophrenia patients and patients with bipolar disorder could be distinguished with an average accuracy of 88%. The model delineating bipolar patients from healthy subjects was less accurate, correctly classifying 67% of the healthy subjects and only 53% of the patients with bipolar disorder. In the latter group, lithium and antipsychotics use had no influence on the classification results. Application of the 1.5 Tesla models on the 3 Tesla validation set yielded average classification accuracies of 76% (healthy vs schizophrenia), 66% (bipolar vs schizophrenia) and 61% (healthy vs bipolar). Most of these works are based on image data. As far as we know, no work has been done on data col-

lected from clinical assessments such as demographic and bipolar disorder features collected from forms and from medical consultations.

In this work, besides working only with annotated clinical data, we also employ a non-propositional machine learning technique, Inductive Logic Programming (ILP), which combines inductive machine learning and logic programming [9]. Compared to other machine learning algorithms, an advantage of ILP is its use of relational logic as a representation language. Learning systems that use this kind of representative language are known as Relational Learning Systems (RLS). In this approach, objects are structurally described, i.e., according to their components and relationships between them. The employment of relational logic as a representative language allows the induction of predicates. As a consequence, the number of concepts that can be learned is expanded. RLS are highly expressive for representing concepts and are able to represent the domain knowledge in such a way that is directly intelligible for people – a fundamental characteristic whenever one aims to extract knowledge from a given domain [10].

In this study, we investigate the use of ILP to generate a multi-relational model for studies of depression relapse in patients with BD. Our study considers demographic and clinical features, organized in multiple tables, which makes the dataset suitable for relational learning [11]. We generate interpretable classifiers, based on first-order logic, that capture the correlation between features included in this study to find patterns of relapse and no-relapse in patients with BD.

Methods

We explored data from a cohort of 139 patients with BD, followed by periods of 6 months to 10 years, from the Bipolar Disorder Program from the Institute of Psychiatry of São Paulo State University outpatient clinic. These data are associated with 102 variables covering psychiatric and medical comorbidity and history of BD, and 109 variables collected during the patient follow-up including medication and symptoms of mood. All participants in the study signed a consent form approved by the Ethics Committee for Research Project Analysis at the Hospital das Clínicas, Faculty of Medicine, University of São Paulo for use of data collected during the follow-up in clinical research.

Initially we identified patients who had continuous periods of remission until depression relapse, or that remained in remission. For this study, we considered only the first depressive relapse of patients. A broader definition of a relapse included both syndromal episodes and subsyndromal clinical states, because reports in the prior literature indicate an association between subsyndromal clinical status of BD and worse clinical and functional outcomes [1, 12, 13]. After filtering, we had 108 patients, where 86 (79.6%) cases are of relapse and 22 (22.7%) are of not relapse, which is representative of the reality of BD patients' outcomes [1 ,3, 4], especially in a university outpatient clinic that treats patients with more severe disorders. All clinical data and follow-up data of the patients were used to generate the relational models. We consider cut-offs of 4 weeks for remission time and classification of mood states based on Hamilton Depression Rating Scale and Young Mania Rating Scale as defined by the International Society of Bipolar Disorder [14, 15, 16].

ILP allows the expert's knowledge to be encoded as background knowledge. We use this characteristic to create predicates that allow searching for relationships between visits of patients. Thus, the ILP system is capable of exploring rules seeking patterns between several visits of the same patient including baseline visit, previous visit to relapse and relapse visit.

Unlike most machine learning approaches, ILP treats its positive and negative training asymmetrically, focusing on inducing rules that match many positive examples and few (ideally zero) negative examples. In this study, we investigate the potential of ILP in two tasks: automatically finding rules that correlate features of (1) relapse and (2) no-relapse cases.

We used Aleph (A Learning Engine for Proposing Hypotheses) [17], an ILP system completely written in Prolog, that was developed at Oxford University. Aleph receives as input: the background knowledge (BK) specified as a logic program; the hypotheses specification (H); an optional set of restrictions (I) relative to acceptable hypotheses; and a finite set of examples $E = E^+ \cup E^-$, where E^+ consists of a "positive" set of examples and E^- a "negative" set of examples. Aleph searches for a hypothesis H, relative to the restrictions I, so that H implies all examples in E^+ and none in E^-. In other words, H is true for all examples in E^+, but it is false for every element of E^-. As a default mode, Aleph applies a covering procedure that constructs a hypothesis through the generation of rules, one by one, using the positive examples. The final outcome consists of a collection of rules (clauses) represented in First Order Logic, known as a theory.

Aleph has several parameters that can be set to guide the search for better rules. We changed the default values of four of these parameters:

clauselength: sets the upper bound on number of literals in an acceptable clause

minpos: sets the minimum number of positive examples that a rule is required to cover.

noise: sets the maximum number of negative examples that a rule is allowed to cover.

nodes: sets the upper bound on the nodes to be explored when searching for an acceptable clause.

We chose 10 for *clauselength* to get rules with the largest number of features and related visits of a patient. We chose 2 for *minpos* in order to produce rules that generalize beyond a single case in the training set at minimum. For *noise* we chose 0, disallowing any rules that misclassify even a single positive example in the training set. Due to the large number of variables, we set the maximum number of *nodes* as 2,000,000.

In our experiments, we use stratified 10-fold cross-validation for evaluating the results. We used the same folds in both experiments (relapse and no-relapse). Metrics presented are collected across all folds, and are related to the test sets. For each experiment, we report the contingency table (TP: True Positives, FN: False Negatives, FP: False Positives and TN: True Negatives), Accuracy (Acc), Recall (Rec), Precision (Prec) and Specificity (Spec). Recall gives the true positive rate, Precision gives the rate of correctly classified positives and Specificity gives the true negative rate.

A statistical analysis of the variables was performed. Sociodemographic and psychiatric characteristics at baseline were compared between relapse and no-relapse groups using the chi-square test for categorical variables (Fisher's exact test if cell counts expected <5), the Student's t-test for normally distributed continuous variables, and the Wilcoxon Mann-Whitney test for non-normally distributed continuous variables. Analyses were conducted using SPSS (V.15.0). All statistical analyses were conducted using a two-sided significance level of α=0.05.

Results

Socio-Demographic and Bipolar Disorder features

In the overall sample, mean age was 40.2 years old (sd. 11.5), ranging from 22 to 76; 72 (66.7%) were female, 84 (77.8%) were Caucasian, 47 (43.5%) were single and 41 (38.0%) married, 16 (14.8%) divorced and 4 (3.7%) widowed, the mean years of education was 11.78 (sd. 3.25) ranging from 3 to 18. The majority of patients, 99 (91.7%), was diagnosed with BD I. Of the 108 subjects evaluated, 86 (79.6%) had a depressive episode relapse. Table 1 shows socio-demographic and BD comparison between groups. No difference was observed regarding the medication in use at baseline: lithium - relapse group (RG) n=29, 33.7% vs. no-relapse group (NRG) n=9, 40.7%, (p=0.53); anticonvulsant - RG n=51, 59.3% vs. NRG n=12, 54.5%, (p=0.69); second generation antipsychotics - RG n=32, 37.2% vs. NRG n=7, 31.8%, (p=0.64); first generation antipsychotics - RG n=3, 4.7% vs. NRG n=0, 0.0%, (p=0.58); antidepressants - RG n=27, 25.0% vs. NRG n=2, 9.1%, (p=0.06).

Table 1– Socio-demographic and Bipolar Disorder characteristics in Relapse and No-Relapse Group

	Relapse	No-relapse	p
n (%)	86 (79.6%)	22 (20.4%)	
Female (n, %)	60 (69.8%)	12 (54.5%)	0.18
Age (mean, sd)	40.9 (12.0)	37.6 (8.9)	0.23
Years of education (mean, sd)	11.8 (3.3)	11.8 (3.0)	0.97
Caucasian (n, %)	69 (80.2%)	15 (68.2%)	0.19
Bipolar Disorder (n, %)	Type I 78 (90.7%) Type II 8 (9.3%)	Type I 21 (95.5%) Type II 1 (4.5%)	0.47
Presence of psychotic symptoms (n, %)	52 (60.5%)	12 (54.5%)	0.53
Age of BD onset (median, iqr)	18.5 (15.0-26.0)	27.0 (17.0-32.0)	0.03
Years with BD (median, iqr)	15.5 (10.0-23.8)	10.0 (6.0-16.0)	0.01
Rapid cycler (n,%)	8 (9.3%)	1 (4.5%)	0.49
Psychiatric comorbidity (n, %)	52 (60.5%)	13 (59.1%)	0.91
Suicide attempt	21 (24.4%)	5 (22.7%)	0.87

Multi-Relational Model

The theories obtained by Aleph were translated from first-order logic to English to make them easier to read. The theory derived for relapse (RG) was composed by four rules and the one derived for no-relapse (NRG) by two rules, which are described next.

Relapse (RG) theory

1. The patient did not have a binge eating disorder, did not have insomnia in the middle of the night in the visit before the relapse, did not have psychotic symptoms and has not been using hypnotic medication.
2. The patient presented loss of interest in activities (hobbies or work), indecision and vacillation in the relapse visit.
3. The patient was not a rapid cycler, was using two different medications, one of them was an anticonvulsant, and at the baseline visit he/she did not have insomnia in the middle of the night and suicidal thoughts.
4. The patient had the BD onset before 17 years old, and did not meet criteria for alcohol abuse lifetime, and at the relapse visit complained about subjective tension and irritability.

No-Relapse (NRG) theory

5. The patient did not present a depressed mood (sadness, hopeless, helpless, worthless) and somatic symptoms.
6. The patient had 13-17 years of education.

The quantitative results from 10-fold cross-validation for each experiment are presented in Table 2. Metrics presented are collected across all folds and are related to the test sets.

Table 2– Accuracy performance

	TP	FN	FP	TN	Acc	Rec	Prec	Spec
RG	79	7	9	13	0.85	0.92	0.90	0.59
NRG	16	6	4	82	0.91	0.73	0.80	0.95

Discussion

Aleph discovered several rules that confirm facts in the literature. Moreover, it discovered patterns that can predict in which conditions a new patient will have the next relapse occurrence, facilitating the immediate detection of the problem and suitable treatment.

As far as we know, this is the first study on the use of ILP to generate a multi-relational clinical model in BD. In addition, this is the first experiment in the major area of psychiatric disorders that explores machine learning classifiers to perform multi-correlation of patients' clinical features based on a longitudinal dataset. In a research area where the finding for interpretable and meaningful patterns is eagerly sought, ILP produced very good initial positive results. The generated rules selected clinical features associated with outcome in BD already described in the scientific literature. Moreover, the associations found by ILP are richer than the ones produced by the so-called black-box models. ILP also produces multivariate correlations as it can combine several variables in the same rule. As ILP is not based on numerical processing, we do not need to worry about missing values.

ILP Theories

BD patients have circadian rhythm genetic abnormalities making them especially sensitive to sleep changes [18]. Sleep disturbance in BD is associated with a more severe course of illness, presence of psychotic symptoms and use of anticonvulsants suicide attempts [19]. All these features were present in the relapse group rules. For example, Rule (1) and Rule (3) mention lack of insomnia, and Rule (3) correlates lack of insomnia and intake of anticonvulsants to a relapse. Anticonvulsants, especially valproic acid, valproate and divalproex are used as mood stabilizers in BD. Better rates of mood stabilization are observed when associated with lithium. When used in monotherapy, they are as effective as lithium [20]. Therefore it is expected that they have a strong role in relapse presence. Lamotrigine, another anticonvulsant with significant positive results in the treatment of BD [21] is included in this group. Anxiety and early age of BD onset are also associated with a poorer outcome [1, 22, 23], and included in Rule (4). BD is strongly associated with suicide. A review study about suicide and BD clinical samples described that between 14-59% of the patients have suicide ideation, 25-56% present at least one suicide attempt during their lifetime and 15-19% die from suicide [24]. Rule (3) is associated with rapid cycling, which was already observed in the literature. Moreover, the use of anticonvulsants in BD is also associated with impulsivity and irritability, both associated with suicide attempts.

Subsyndromal depressive symptoms are strongly associated with depressive relapse [13]. In all rules there is at least one depressive symptom. Interestingly, the rules also describe the absence of specific depressive symptoms as a predictor of depressive relapse. Rule (2) consists only of depressive symptoms. In the No-relapse theory, Rule (5) consists only of depressive symptoms, strengthening once more the power of subsyndromal role in depressive relapse. Rule (6) is associated with years of education. Education is associated with higher levels of resilience, the ability to cope with stress or adversity, which is positively associated with better outcomes [25].

A possible limitation concerning our sample is the difference of BD age onset and years with BD between groups. The early BD onset as a factor associated with relapse was observed in the statistical and ILP analyses. It is well known in the literature the negative impact of early age of onset in BD outcome [1]. Other limitations could be related to the follow-up and number of visits per subject, once they were naturalistically followed and the visits were scheduled according to mood changes. A structured follow-up program could reduce this bias in future machine learning studies.

ILP as a Classifier

Besides producing rules that can represent patient patterns and allowing for the representation of new background knowledge that can be incorporated in to the raw data, ILP rules can also work as classifiers that can predict relapse or no-relapse of future cases. Results in Table 2 show the expected performance of a classifier based on rules produced by Aleph on our dataset. As explained before, these results were generated using 10-fold cross-validation (which is a standard machine learning validation procedure) and results are reported for the test set.

The first row of Table 2 shows results for rules that can predict in which situation a patient will have a relapse. These rules are based on the four rules discussed in the Results section. This classifier has a very good accuracy of 85%, being able to correctly predict 92 out of the 108 patients. On the relapse class, it does quite well with a 92% Recall, meaning that it misclassifies only 7 patients that actually have a relapse. This has a very important clinical meaning, since psychiatrists can concentrate their attention on the most needed patients with a very regular follow-up, while scheduling other patients to a more spaced time interval. It is important to note that the current practice routine considers all patients as potential relapse cases.

The second row of Table 2 shows similar results, but from the point of view of the no-relapse patients. If we look at the error rate of this classifier, it can perform even better than the first classifier for relapse cases. In other words, if the classifier says that a case is no-relapse, only 4 actual relapse cases are misclassified, while most of the no-relapse cases (16 out of 22, 73%) are still correctly classified. This means that 73% of the actual no-relapse cases can have a different treatment and have less regular visits to the psychiatrist, again leaving more time for the psychiatrists to concentrate on the actual relapse cases.

These are very encouraging results which open a new path to novel discoveries in the area of mental disorders.

Conclusion

In this work, we use ILP to develop a multi-relational model that highlights patterns of relapse and non-relapse in patients with BD. The results confirm that ILP can generate accurate relational models. The models are easy to interpret, since they are based on a rule-based language helping the experts to better understand the found patterns. By using this language, the expert can also add new knowledge to the raw data. Rules pinpoint features that were already reported in the literature such as use of convulsants and cases of insomnia, among others, but more importantly, they correlate many variables, contrary to most studies in the literature that only find uni- or bivariate correlations.

We used 10-fold cross-validation on a set of 108 patients, 22 with no-relapse and 86 with relapse. Results show that we can achieve very good accuracy when predicting either class and very good Recall (up to 95% specificity on the no-relapse class), numbers that are better than the ones found in the literature so far.

Another advantage of ILP is that it can handle multiple tables, which is not so trivial with propositional methods that require that the data is represented in a flat table format. For example, in our experimental setting, each patient can have a variable number of visits per month (up to 45 non-consecutive visits in 125 months), and each visit is associated to several different characteristics (totaling 211 variables per patient). Generating a flat table for this kind of data keeping the absolute time of the visit would require a huge, sparse, and redundant table with 13,727 columns for each patient. In addition, models produced with the ILP learner would not be comparable with models produced with this flat table.

Some ILP limitations concern the rapid growth of the search space relative to database growth. Such situations may lead to a high computational cost for the rule construction, thus making it more difficult to find the best solution.

ILP and other approaches in machine learning could be especially useful in exploratory studies with a multifactorial approach, as translational research, and the possibility to find patterns with impressive precision. The mental disorder field could be explored as a tool to link neurobiology with clinical phenotype. More research in other areas of mental disorders must be done to fully uncover the ILP potential.

Acknowledgments

Rogerio Salvini and Rodrigo Dias would like to thank Daniele Madureira from the National Laboratory for Scientific Computing (LNCC, Petropolis, Brazil) for discussions and support. Rogerio Salvini and Inês Dutra would like to acknowledge the FCT project "ABLe: Advice-Based Learning for Health Care" (PTDC/ EEI-SII/ 2094/ 2012). This research was carried out without funding.

References

[1] Goodwin FK, Jamison KR, Ghaemi SN. Manic-depressive illness : bipolar disorders and recurrent depression. 2nd ed. New York, N.Y.: Oxford University Press; 2007.

[2] World Health Organization, The global burdem of disease: 2004 update. World Health Organization Press; 2004. pp. 46.

[3] Gitlin MJ, Swendsen J, Heller TI, Hammen C. Relapse and impairment in bipolar disorder. American Journal of Psychiatry, 1995; 152:1635-40.

[4] Tohen M, Waternaux GM, Tsuang MT. Outcome in mania: a 4-year prospective follow-up of 75 patients utilising survival analysis. Archives of General Psychiatry, 1990; 47:1106–1111.

[5] Oquendo MA, Baca-Garcia E, Artés-Rodríguez A, Perez-Cruz F, Galfalvy HC, Blasco-Fontecilla H, Madigan D, Duan N. Machine learning and data mining: strategies for

hypothesis generation. Molecular Psychiatry, 2012; 17:956–959.
[6] Phillips ML, Kupfer DJ. Bipolar disorder diagnosis: challenges and future directions. Lancet. 2013; 381(9878):1663-71.
[7] Mourão-Miranda J, Oliveira L, Ladouceur CD, Marquand A, Brammer M, Birmaher B, Axelson D, Phillips ML. Pattern recognition and functional neuroimaging help to discriminate healthy adolescents at risk for mood disorders from low risk adolescents. PLoS One. 2012; 7(2).
[8] Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, Hulshoff Pol HE, Kahn RS. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. Neuroimage. 2014; 84:299-306.
[9] Muggleton S. Inductive logic programming. New Generation Computing, 1991; 8(4):295-318.
[10] Michie D. Machine learning in the next five years. In Proc. Third European Working Session on Learning, Pitman, London, 1988; pp.107-122.
[11] Raedt LD. Logical and Relational Learning. Springer, 2008.
[12] Marangell LB, Dennehy EB, Miyahara S, et al. The functional impact of subsyndromal depressive symptoms in bipolar disorder: data from STEP-BD. J Affect Disord. 2009; 114(1-3):58-67.
[13] Tohen M, Bowden CL, Calabrese JR. Influence of sub-syndromal symptoms after remission from manic or mixed episodes. Br J Psychiatry. 2006; 189:515-519.
[14] Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, Malhi GS, Calabrese JR, Nolen WA, Vieta E, Kapczinski F, Goodwin GM, Suppes T, Sachs GS, Chengappa KR, Grunze H, Mitchell PB, Kanba S, Berk M. The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. Bipolar Disord. 2009; 11(5):453-73.
[15] Cetin T, Guloksuz S, Cetin EA, Gazioglu SB, Deniz G, Oral ET, van Os J. Plasma concentrations of soluble cytokine receptors in euthymic bipolar patients with and without subsyndromal symptoms. BMC Psychiatry. 2012; 12:158.
[16] Kulkarni J, Filia S, Berk L, Filia K, Dodd S, de Castella A, Brnabic AJ, Lowry AJ, Kelin K, Montgomery W, Fitzgerald PB, Berk M. Treatment and outcomes of an Australian cohort of outpatients with bipolar I or schizoaffective disorder over twenty-four months: implications for clinical practice. BMC Psychiatry. 2012; 12:228.
[17] Srinivasan A. The Aleph Manual, 2001.
[18] Murray G, Harvey A. Circadian rhythms and sleep in bipolar disorder. Bipolar Disord. 2010; 12(5):459-72.
[19] Sylvia LG, Dupuy JM, Ostacher MJ, Cowperthwait CM, Hay AC, Sachs GS, Nierenberg AA, Perlis RH. Sleep Disturbance in Euthymic Bipolar Patients. J Psychopharmacol. 2012; 26(8):1108-12.
[20] Cipriani A, Reid K, Young AH, Macritchie K, Geddes J. Valproic acid, valproate and divalproex in the maintenance treatment of bipolar disorder. Cochrane Database Syst Rev. 2013; 10:CD003196.
[21] Popovic D, Reinares M, Amann B, Salamero M, Vieta E. Number needed to treat analyses of drugs used for maintenance treatment of bipolar disorder. Psychopharmacology (Berl). 2011; 213(4):657-67.
[22] Otto MW, Simon NM, Wisniewski SR, Miklowitz DJ, Kogan JN, Reilly-Harrington NA, Frank E, Nierenberg AA, Marangell LB, Sagduyu K, Weiss RD, Miyahara S, Thas ME, Sachs GS, Pollack MH;STEP-BD Investigators. Prospective 12-month course of bipolar disorder in outpatients with and without comorbid anxiety disorders.Br J Psychiatry. 2006; 189:20-5.
[23] Bowden CL, Perlis RH, Thase ME, Ketter TA, Ostacher MM, Calabrese JR, Reilly-Harrington NA, Gonzalez JM, Singh V, Nierenberg AA, Sachs GS. Aims and results of the NIMH systematic treatment enhancement program for bipolar disorder (STEP-BD; CNS Neurosci Ther. 2012; 18(3):243-9.
[24] Abreu LN, Lafer B, Baca-Garcia E, et al. Suicidal ideation and suicide attempts in bipolar disorder type 1: an update for the clin ician. Rev Bras Psiquiatr. 2009; 31(3):271–280.
[25] Choi JW, Cha B, Jang J, Park CS, Kim BJ, Lee CS, Lee SJ. Resilience and impulsivity in euthymic patients with bipolar disorder.J Affect Disord. 2015; 170:172-7.

Address for correspondence

Rogerio Salvini

Instituto de Informática, Universidade Federal de Goiás (INF/UFG)
E-mail: rogeriosalvini@inf.ufg.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-746

Thermal Signal Analysis for Breast Cancer Risk Verification

Lincoln F. Silva[a], Giomar O. Sequeiros[a], Maria Lúcia O. Santos[b], Cristina A. P. Fontes[b], Débora C. Muchaluat-Saade[a] and Aura Conci[a]

[a] *Institute of Computing, Fluminense Federal University, Niterói, RJ, Brazil*
[b] *Department of Radiology, Fluminense Federal University, Niterói, RJ, Brazil*

Abstract

Breast cancer is the second most common cancer in the world. Currently, there are no effective methods to prevent this disease. However, early diagnosis increases chances of remission. Breast thermography is an option to be considered in screening strategies. This paper proposes a new dynamic breast thermography analysis technique in order to identify patients at risk for breast cancer. Thermal signals from patients of the Antonio Pedro University Hospital (HUAP), available at the Mastology Database for Research with Infrared Image - DMR-IR were used to validate the study. First, each patient´s images are registered. Then, the breast region is divided into subregions of 3x3 pixels and the average temperature from each of these regions is observed in all images of the same patient. Features of the thermal signals of such subregions are calculated. Then, the k-means algorithm is applied over feature vectors building two clusters. Silhouette index, Davies-Bouldin index and Calinski-Harabasz index are applied to evaluate the clustering. The test results showed that the methodology presented in this paper is able to identify patients with breast cancer. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved.

Keywords:

Breast cancer; Thermal signals; Clustering; Validity indexes; Classification.

Introduction

Breast cancer is the most common cancer among women worldwide. However, when diagnosed and treated in early stages, this cancer type has relatively good prognosis [1].

Screening is a strategy adopted by health authorities in order to identify women who are at initial stages of breast disease. Thus, it is necessary to develop methods and techniques in order to improve screening procedures because even mammography, which is considered the gold standard for cancer detection, has its limitations, such as high false positive classification rate, insufficient effectiveness in dense breasts and use of ionizing radiation to form breast images [2][3].

Since the cancerous tissue temperature is generally higher than healthy surrounding tissues, thermography has been considered a promising screening method for breast cancer detection by generating images that reveal the heat distribution on the breast surface [3].

The thermal signals used in this work originated from dynamic thermography, which is a method for monitoring the dynamic response of the skin temperature after thermal stimulus. In other methods for detecting breast cancer, the thermal stimulus most utilized is the application of air flow directed to the breasts by an electric fan [4][5][6]. The cooling of the breasts, theoretically, improves the thermal contrast between healthy and unhealthy tissues in the image, because blood vessels promoted by cancerous tumors do not have muscular layer and neural regulation like embryological vessels. These vessels are only endothelial tubes and therefore do not contract in response to sympathetic stimulation. For that reason breast region with cancerous tumors remain with unchanged temperature while the healthy part of the breast is cooled down [7].

When compared to static thermography, dynamic thermography is faster and more robust. Indeed, static thermography requires rigid environmental conditions and significantly long time for acclimatization of the patient to examining room conditions. On the other hand, dynamic thermography is much less dependent of the conditions and temperature of the examining room [8].

This paper proposes a new dynamic breast thermography analysis technique in order to identify patients at risk for breast cancer. Dynamic thermal signals from patients of the HUAP, available at the Mastology Database for Research with Infrared Image - DMR-IR [9] were used for validating this hypothesis. First, each patient has her images registered. Second, the breast region is divided into regions of 3x3 pixels and average temperature from each of these regions is observed in all images of the same patient. Features of the thermal signals of such regions are calculated. Then, the *k*-means algorithm is applied over the feature vectors building two clusters. Silhouette, Davies-Bouldin and Calinski-Harabasz indexes are applied to evaluate clustering. The test results demonstrated that the methodology presented in this paper is able to identify these patients. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved.

The remainder of the paper is organized as follows: related work section contains main work identified in the literature; the methodology proposed in this work is detailed in the section of same name; in the section tests and results, results are presented and discussed; and the conclusion section concludes this paper and indicates future work.

Related work

In Gerasimova *et al.* [10], dynamic thermography has been performed in 46 histopathologically proven breast malignant and benign tumor cases before surgery. From regions with and without tumor, thermal signals have been generated and Fourier Analysis, Wavelet Transform and phase diagram have been applied in order to examine the behavior of these signals. The authors have concluded that chaotic phase diagrams correspond to the healthy tissue; while for cancerous tissues, irregular shape in phase space is typical. Furthermore, results indicate that for healthy tissue, the thermal signals are anti-

correlated, whereas in cancerous tissue, thermal noise correlation has been observed. According to these authors, these results indicate that abnormal tissue has no ability to adapt to external influences. In addition, according to the authors, the results are consistent to the golden rule in biomedical sciences. This rule states that healthy and normal biomedical systems are often very complex and that the complexity decreases when an abnormality or disease occurs.

In another work, Gerasimova *et al.* [3] have performed multifractal analysis over breast thermal signals produced by dynamic thermography to detect differences in behavior between tissue with malignant tumor and healthy tissue. The modulus maxima wavelet transform method has been applied to characterize the multifractal properties of the one-dimensional thermal signals of cancerous and healthy breasts. The authors have concluded that complex scalar multifractal properties of the signals over autonomic regulation are drastically altered when the disease is present. In such a study, both breasts of 9 women were imaged, 6 with cancer and 3 healthy. A photovoltaic detector camera *InSb* was used. During acquisition of the images, the patient remains seated with the arms down in order to avoid discomfort. Frontal images were captured at 1 meter distance in a controlled temperature environment between 20ºC to 22ºC. Each set contains 30,000 image frames acquired during 10 minutes. Skin surface markers were used as landmarks for image registration phase in order to eliminate motion artifacts and do not hinder the analysis. In sick patients, the tumor region and the symmetrically positioned region in the other breast are delimited by square regions of 8 x 8 pixels. The analysis has been performed only within these regions. The methodology is able to discriminate between healthy region and region with tumor. Healthy regions exhibited signals with multifractal dimension as complexity signature and tumor regions exhibited signs with monofractal dimension as evidence of loss of complexity.

Recently, Gerasimova *et al.* published another work [11] using a much larger database, with 33 patients with histopathologically proven cancer and 14 healthy volunteers for control. The findings reaffirm the results of the prior study [3].

Scully *et al.*[12] also conducted analysis of skin thermal signal in patients with skin cancer. The temperature monitoring was performed by dynamic thermography. 1,500 thermograms, containing the region with disease, were collected using a camera FLIR SC7700, during 25 minutes. The collected images have been registered in relation to the first image and Wavelet Transform was used to analyze the thermal signals. They found significant differences between the region of control and tumor regions.

Herman [8] claims that dynamic thermography is able to measure the difference in infrared emission between healthy tissue and melanoma during the temperature recovery process after removal from cold stress. Test results suggest that the temperature is higher in cancerous lesion than in non-cancerous lesion during the first 45-60 seconds of thermal recovery. In this methodology, the region with tumor is surrounded by a bounding rectangle. Then, a digital photograph and an infrared image are captured. The infrared image observes the situation of steady state under ambient conditions. After, the skin is cooled by cold airflow per 1 minute. Next, infrared images are captured during 200 seconds. The methodology was applied in 37 patients, 3 with histopathologically confirmed cancerous lesions. The method achieved 100% accuracy.

Liu *et al.* [13] used thermograms to observe the forearm temperature variation and perform classification between three tissues: with micro-vascularity, with large veins and no veins. 3,000 thermographs were captured during 12 minutes and were registered afterwards with reference to the first image of the sequence. Thermal signals of each pixel were built, totaling 81,000 signs. Clustering using *k*-means algorithm and *Short Time Series (STS)* distance were applied over the signs to identify tissues with similar behavior in time. The authors demonstrated differences between three tissue types in terms of temperature change over time (via temporal profiles), magnitude of frequency response (via FFT), and coherency (via wavelet phase coherence and power spectrum correlation).

Unlike previous work, our methodology applies machine learning to decide whether or not a given patient is at risk of breast cancer.

Methodology

Our methodology steps are: image registration; thermal signal construction; thermal signal feature extraction; clustering; clustering evaluation and classification model building.

Image Registration

During examination by dynamic thermography, the patient may perform involuntary movements from breathing and momentary imbalance. These movements cause differences from one image to another and consequently generate thermal noise in the formed signs. Figure 1 shows (a) the first image of a particular patient, (b) the seventeenth image of the same patient, and (c) the result of the subtraction of (b) and (a) images. It is possible to see the difference between the images by observing image (c) which illustrates the effect of the movement of the patient during the examination.

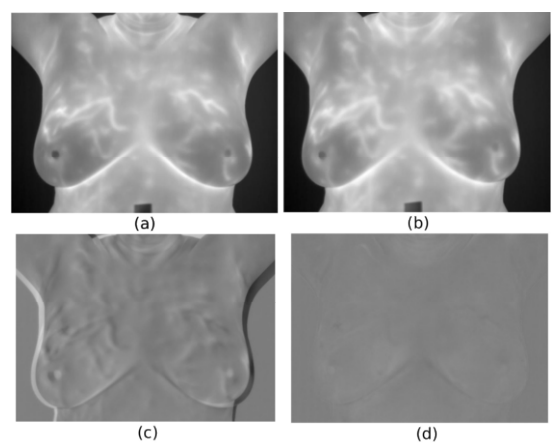


Figure 1- Image registration: (a) first image of sequence; (b) seventeenth image of sequence; (c) subtraction of (b) by (a) before registration; (d) the same subtraction after registration of these images.

In order to decrease the effects of these movements on thermal signals, registration of all images is performed. Registration is a process in which images, roughly speaking, are "matched". Considering two images (reference and sensitive) of the same scene, registration seeks to create a relationship between them to achieve the best overlap possible [14]. In this work, the first image of the sequence was considered the reference (immutable) image and the remaining images have been considered sensitive (transformable) images.

The registration is performed in two steps. In the first stage, a rigid body registration is applied based on intensities with geometric transformation consisting of translation, rotation, and scale. The second step uses non-rigid 2D registration with Residual Complexity (RC) [15].

The result of the registration can be seen in Figure 1(d). That image represents the subtraction of image (b) by image (a). After registration, the difference is much smaller than that one shown in image (c), before registration.

It is noteworthy that the data used in the analysis described in this paper are the temperature matrices of the breast surface, captured by an infrared camera. Thus the temperature matrix is converted to an image in gray tones, then the registration is performed in this image and the transformations are transferred back to the temperature matrix.

Thermal Signal Construction

After registration of the images, the formation of the thermal signal from the dynamic sequence of 20 images of a particular patient comprises the following steps:

1. The region of the breasts of the first image of the sequence is segmented creating a mask as shown in Figure 2(a);
2. The image region corresponding to the created mask is divided into a grid of 3x3 pixel squares R_k(Figure 2(b)), with k=1,...,p , where *p* is the amount of formed squares;
3. The averaged temperature of each square R_k is observed in all twenty images of the sequence producing the $S_k = (t_1,t_2,...,t_{20})$ thermal signal;
4. These twenty points are interpolated by cubic convolution yielding a signal of 1,901 points (in each interval between two points, 100 new points have been inserted, then 100x19 intervals = 1,900 + 1 (last point) = 1,901).

The S_k series values are ordered chronologically, *i.e.,* $t_1$is the mean temperature square R_k in the first image of the sequence, t_2 is the mean temperature square R_k in the second image of the sequence, and so on.

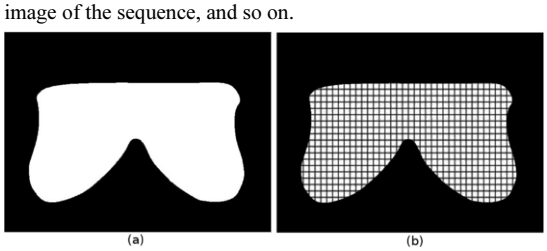


Figure 2 – Mask and grid of squares

Figure 3 shows the thermal signals of a sick patient and Figure 4 shows the thermal signals of a healthy patient. The *x*-axis represents time (the 20 points of each serie). The *y*-axis contains each of S_k (the index) of each thermal signal. The *z*-axis indicates the temperature in Celsius degrees. It can be seen that there is a group of thermal signals with higher temperature (red color) and with sharper increase in the early stages of the temperature recovery after thermal stress for the sick patient (Figure 3) however the same is not true for the healthy patient (Figure 4). The extracted features attempt to emphasize these differences.

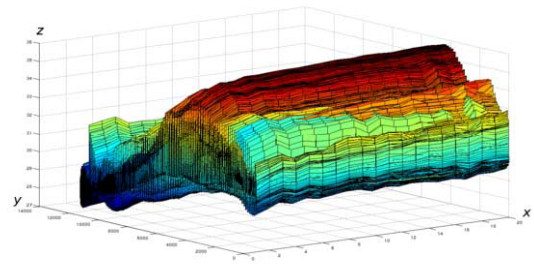


Figure 3 - Thermal signs of a sick patient.

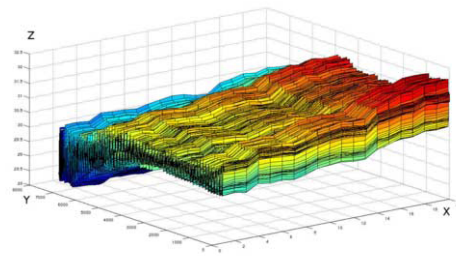


Figure 4 - Thermal signs of a healthy patient.

Thermal Signal Feature Extraction

The S_k time series can be treated as a biological signal of temperature t_i, i=1 ... N, where N=20. In biomedical systems, the evaluation of the complexity (main feature) is an important factor for diagnosis. This is possible through identifying and evaluating the complexity of the signal created by these systems [16].

In this work we used two signal complexity measures as features: the signal mobility which uses normalized first-order variations of the signal; and the signal complexity that uses second-order variations of the signal. The two characteristics measure the degree of variation along the signal. [16]. Considering a particular patient, from each sign t_i first-order variations are calculated, defined by d_i=t$_i$ − t_{i-1}, i=2 ... N, and second-order variations, defined by g_i=d$_i$ − d_{i-1}, i=3 ... N.

$$S_0 = \sqrt{\frac{\sum_{i=1}^{N} t_i^2}{N}} \quad (1)$$

$$S_1 = \sqrt{\frac{\sum_{i=2}^{N} d_i^2}{N-1}} \quad (2)$$

$$S_2 = \sqrt{\frac{\sum_{i=3}^{N} g_i^2}{N-2}} \quad (3)$$

Measures of complexity and mobility are defined in Equations (4) and (5), respectively:

$$C = \sqrt{\frac{S_2^2}{S_1^2} - \frac{S_1^2}{S_0^2}} \quad (4) \qquad M = \frac{S_1}{S_0} \quad (5)$$

Clustering, Evaluation of Clusters and Classification

The calculated features for each signal are stored in a vector. Clustering is performed over the set formed by feature vectors. In this work, we applied the *k*-means algorithm for clustering the vector set into two clusters. The two expected clusters are: thermal signals generated from diseased tissue and thermal signals generated from healthy tissue. We believe that for healthy patients the two clusters formed are very similar (healthy patients have no diseased tissues and, therefore, all the signs of both breasts are very similar), whereas for sick

patients the two clusters formed are more compact and well separated. The *k*-means algorithm has been performed by applying the correlation distance and three repeated steps, where each iteration has a different set of initial cluster centroid positions. The formed clusters are evaluated by clustering validity indexes. In this work, we applied: Silhouette index [17], Davies-Bouldin [18], Krzanowski-Lai index [19] and Calinski-Harabasz index [20].

The Silhouette index indicates the number of clusters that best separate the data set, *i.e.*, the maximum value of this index indicates the optimal number of compact and well separated clusters. It is defined by:

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (6)$$

where $a_i$is the average dissimilarity (*i.e.* how far away two elements are from each other using, *e.g.*, Euclidean distance) of i with all other data within the same cluster, b_i is the lowest average dissimilarity of i to any other cluster of which i is not a member. Note that $-1 \ S \ 1$.

The Calinski-Harabasz index is defined in terms of the traces of the between-clusters and within-cluster scatter matrices. The trace is defined to be the sum of the elements on the main diagonal of the matrix. Calculated for each possible cluster solution, the maximal achieved index value indicates the best data clustering. It is calculated using the following equation:

$$\mathrm{CH}=\frac{[\mathrm{traceB}/K-1]}{[\mathrm{traceW}/N-k]} \text{for } \mathrm{K} \in \mathbb{N} \quad (7)$$

where B denotes the error sum of squares between different clusters (inter-cluster) and W the squared differences of all objects in a cluster from their respective cluster center (intra-cluster).

The Krzanowski-Lai is based on the square differences of all feature vectors in a cluster from their respective cluster center. It is calculated using the following equation:

$$\mathrm{KL}(K) = \left|\frac{\mathrm{DIFF}(K)}{\mathrm{DIFF(K+1)}}\right| \quad (8)$$

where

$$\mathrm{DIFF}(K) = (K-1)^{\frac{J}{2}} \cdot \mathrm{traceW}_{K-1} - K^{\frac{2}{J}} \cdot \mathrm{traceW}_K \quad (9)$$

which is the difference between a clustering in K and a clustering in $K-1$ clusters. J is the number of variables that have been measured on each $x_i \in X$and *trace* W_K the sum of square function that corresponds to the clustering in K clusters.

Regarding the Davies-Bouldin index, it requires the dispersion measure and the cluster similarity (how much the elements resemble each other) measure. Thus it is defined as the ratio of the dispersion within the clusters and the separation between clusters and is calculated by:

$$\mathrm{DB}(K) = \frac{1}{K}\sum_{\mathrm{k}=1}^{K} R_k \text{ ,K} \in \mathbb{N} \quad (10)$$

where

$$\left(\frac{S_k+S_J}{d_{\mathrm{k,j}}}\right), \ \mathrm{k} \in [1,\ldots,\mathrm{K}] \quad (11)$$

and

$$S_k = \frac{1}{\sum w_{\mathrm{k,i}}}\sum_{\mathrm{i}=1}^{n}\|x_i - \bar{x}_k\| \text{ ,k} \in [1,\ldots,\mathrm{K}] \quad (12)$$

as well as

$$d_{\mathrm{k,j}} = \left\|\bar{x}_k - \bar{x}_j\right\| \quad (13)$$

In the case of this index, the minimum observed value indicates the best solution for clustering.

Clustering validity index values have been used as features to generate classification models. These values have been submitted to Weka tool [22] applying the *MultilayerPerceptron* (a neural network), the BayesNet (a bayesian network), and the J48 (a decision tree) algorithms.

Tests and results

The used infrared images are from the *Mastology Database for Research with Infrared Image - DMR-IR*. This database is described by Silva *et al.* [9], where the image acquisition protocol is also described with more details. Briefly, in the execution of the protocol, regions of the breasts and armpits of the patient are cooled by an electric fan for some minutes in an environment with controlled temperature (20ºC to 22ºC). After cooling, 20 images are captured during 5 minutes. Images are captured using a FLIR thermal camera, model SC620 [21]. The sensitivity of the camera is smaller than 0.04ºC, the detectable temperature range is between -40ºC and 500ºC and images are generated with dimension of 640x480 pixels.

Tests have been performed with images of 22 patients including 11 with histopathologically proven cancer and 11 healthy. For each patient, all images obtained by dynamic thermography are submitted to the steps of the methodology described in the previous section and implemented in MatLab. The results are in Table 1, Table 2 and Table 3, respectively. In all tests, the *k*-Folds Cross Validation technique has been applied in order to evaluate the classification results, with *k=3*. In all tables, TP Rate is the true positive rate, FP is the false positive rate and AUC is the area under the ROC.

Table 1: Neural network classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.81	0.00	1.00	1.00
Healthy	1.00	0.18	0.85	1.00
Weighted Avg	0.91	0.09	0.93	1.00

Confusion Matrix	Expected class	
	With cancer	Healthy
Classified as with cancer	9	2
Classified as healthy	0	11
Correctly Classified Instances	20	90.91%

Table 2: Bayesian network classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.73	0.00	1.00	0.93
Healthy	1.00	0.27	0.79	0.93
Weighted Avg	0.86	0.14	0.89	0.93

Confusion Matrix	Expected class	
	With cancer	Healthy
Classified as with cancer	8	3
Classified as healthy	0	11
Correctly Classified Instances	19	86.36%

Table 3: Decision tree classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.91	0.09	0.91	0.92
Healthy	0.91	0.09	0.91	0.92
Weighted Avg	0.91	0.09	0.91	0.92

Confusion Matrix	Expected class	
	With cancer	Healthy
Classified as with cancer	10	1
Classified as healthy	1	10
Correctly Classified Instances	20	90.91%

The percentage of correctly classified instances ranges from 86.36% to 91.90% considering the results of all classifiers. From the results obtained, we can conclude that the proposed methodology is potentially able to identify patients at risk for breast cancer. However, we will definitely confirm the effectiveness of the methodology when we extend our test database. At the moment, we are working hard to that effect.

Conclusion

Breast cancer has been killing many women around the world. This work has exploited the fact that breast regions with cancer produce thermal signals with complexity alteration when analyzed over time. Thermal signals have been generated by dynamic thermography. The complexity of these signals has been measured by well-known measures described in the literature. Then the *k*-means algorithm has been applied to the extracted features for clustering them into two clusters. Clustering validity indexes were applied to identify patients at risk for breast cancer. The test results demonstrated that the methodology presented in this paper is able to identify these patients. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved. In future work, other features will be extracted as well as other clustering algorithms and clustering validity indexes will be tested.

Acknowledgements

Thanks to Brazilian CAPES, CNPq and FAPERJ agencies for partially funding this work and to FAPEMA for funding the camera.

References

[1] M. H. Forouzanfar, K. J. Foreman, A. M. Delossantos, R. Lozano, A. D. Lopez, C. J. L. Murray, and M. Naghavi, Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis, *The Lancet*, 378 (9801), 1461–1484, 2011.

[2] C. Bouchardy; G. Fioretta; H. M. Verkooijen; et al. Recent increase of breast cancer incidence among women under the age of forty, *British Journal of Cancer*; 96 (11), 1743-1746, 2007.

[3] E. Gerasimova; B. Audit; S. G. Roux; A. Khalil; F. Argoul; O. Naimark1 and A. Arneodo, *Multifractal analysis of dynamic infrared imaging of breast cancer, EPL (Europhysics Letters)*, 104 (6), 2013.

[4] P. Kapoor and S. Prasad, Image processing for early diagnosis of breast cancer using infrared images, *2nd Int. Conf. on Computer and Automation Engineering*, 13 (1), 564-566, 2010.

[5] J. Koay; C. Herry and M. Frize, Analysis of breast thermography with an artificial neural network, *Engineering in Medicineand Biology Society – IEMBS*, 1 (1), 1159-1162, 2004.

[6] Y. Ohashi and I Uchida, Applying dynamic thermography in the diagnosis of breast cancer, *Engineering in Medicine and Biology Magazine, IEEE*, 19 (3), 42-51, 2000.

[7] W. C. Amalu, Nondestructive Testing of the Human Breast: The Validity of Dynamic Stress Testing in Medical Infrared Breast Imaging, *Engineering in Medicine and Biology Society*, 1, 1174-1177, 2004.

[8] C. Herman, The role of dynamic infrared imaging in melanoma diagnosis, *Expert Rev Dermatol*, 8 (2), 177-184, 2013.

[9] L. F. Silva; D. C. M. Saade; G. O. Sequeiros; A. C. Silva; A. C. Paiva; R. S. Bravo and A. Conci, A New Database for Breast Research with Infrared Image, *Journal of Medical Imaging and Health Informatics*, 4 (1), 2014.

[10] E. Gerasimova; O. Plekhov; Yu.Bayandin; O.Naimark and G. Freynd, Identification of breast cancer using analysis of thermal signals by nonlinear dynamics methods, *International Conference on Quantitative InfraRed Thermography*, 2012.

[11] E. Gerasimova; B. Audit; S. Roux; A. Khalil; O. Gileva; F. Argoul; O. Naimark and A. Arneodo, Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis, *Frontiers in Physiology*, 5 (176), 2014,

[12] C. G. Scully; W. Liu; J. Meyer; A. Dementyev; K.H. Chon; P. Innominato; F. Lévi and A.M. Gorbach, Time-Frequency Analysis of Skin Temperature in a Patient with a Surface Tumor Monitored with Infrared Imaging, *International Conference on Quantitative InfraRed Thermography*, 81-88, 2010.

[13] W. Liu; J. Meyer; C.G. Scully; E. Elster and A. M. Gorbach, Observing Temperature Fluctuations in Humans Using Infrared Imaging, *10th International Conference on Quantitative InfraRed Thermography*, 55-64, 2010.

[14] A. Ardeshir Goshtasby, 2-D and 3-D Image Registration: for Medical, *Remote Sensing, and Industrial Applications, Wiley, New Jersey*, ISBN 0-471-64954, 2005

[15] A. Myronenko and S. Xubo, "Intensity-Based Image Registration by Minimizing Residual Complexity," *Medical Imaging, IEEE Transactions on*, 29 (11), 1882-1891, Nov. 2010.

[16] N. Kayvan and R. Splinter, Biomedical Signal and Image Processing, *CRC Press, edition 1th*, 2005.

[17] P. J. Rousseeuw, Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Computational and Applied Mathematics*, 20, 53–65, 1987.

[18] D. L. Davies and D. W. Bouldin, A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1 (2), pp. 224–227, 1979.

[19] W. Krzanowski and Y. Lai, A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44 (1), pp. 23–34, 1988.

[20] T. Calinski and J. Harabasz, A Dendrite Method for Cluster Analysis, *Communications in Statistics – Theory and Methods*, 3 (1), pp. 1-27, 1974.

[21] Flir, http://www.flir.com/BR/, Accessed on April 2014, 1999.

[22] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten; The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11 (1), 2009.

Address for correspondence

Lincoln F. Silva: lsilva@ic.uff.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-751

Time-Series Data Analysis of Long-Term Home Blood Pressure Measurements in Relation to Lifestyle

Hiroshi Takeuchi, Naoki Kodama, Shingo Takahashi

Department of Healthcare Informatics, Takasaki University of Health and Welfare, Takasaki, Gunma, Japan

Abstract

We conducted a long-term time-series analysis of an individual's home blood pressure measurements, stored on a personal healthcare system in cloud, relative to the individual's lifestyle. In addition to daily scattering, apparent seasonal variations were observed in both systolic and diastolic blood pressure measurements. We examined the effect of seasonal variations on the outcome of a healthcare data mining process that extracts rules between blood pressure measurements and lifestyle components such as exercise and diet, and found that the daily blood pressure data approached a normal distribution when adjusted for the seasonal variations. This implies that an adjustment is desirable in order to produce appropriate rules in the healthcare data mining process.

Keywords:

Personal Healthcare; Home Blood Pressure; Time-Series Data Analysis; Healthcare Data Mining.

Introduction

The mHealth and pHealth are currently the topics of great interest [1,2]. In particular, the demand for personalized healthcare systems to prevent diseases and improve health has been increasing [3]. Within this context, we have developed a personal dynamic healthcare system (PDHS) using cloud computing [4]. This system enables storage of the personal health and lifestyle time-series data in a database using a mobile device, and allows for the extraction of useful information, such as the rules and patterns of the lifestyles and health conditions contained within that data. We call this 'healthcare data mining.'

In the healthcare data mining process [5], we first check the correlation between variations of the time-series health data and summaries of the overall time-series lifestyle data. If the correlation coefficient is larger than a certain threshold, the lifestyle is selected as an independent variable in the data mining process relevant to the health condition. In this study, we analyzed the long-term time-series data of an individual's home blood pressure measurements in relation to his daily energy expenditure and supply (i.e. food-intake). Our specific focus is the effect of the seasonal variations in the home blood pressure measurements on the data mining process that extracts rules between blood pressure and energy expenditure/supply.

Materials and Methods

Subject and Method of Data Acquisition

The subject of this study was a 57-year-old (when the data storage began) male, who lives in Tokyo, Japan, and goes to the office on weekdays.

The blood pressure measurements (systolic and diastolic) were taken at home every morning upon waking, with an automatic blood pressure meter (Omron, Japan) using an oscillometric method. Each measurement was performed three times and the average value was then recorded. Energy expenditure due to exercise was measured with a wearable monitor (Omron, Japan) and the energy supply was estimated from each day's breakfast, lunch, and dinner contents. These data were entered via a mobile device and stored in a database using PDHS from 2004/06/01 to 2012/12/31.

Healthcare Data Mining Algorithm

Time-Delayed Correlation Analysis

The time-series data analysis described here is based on the simple idea that the accumulated effects of lifestyle, such as exercise and the diet, could affect personal health with some delay [5]. This delay may reflect complex bio-reactions such as those of the metabolism in a human body. In our analysis, the accumulation of lifestyle effects is represented by a summary of lifestyle data, namely energy supply and expenditure. The accumulated effects may cause corresponding variations in the vital signs (e.g. blood pressure), representing health conditions, with some delay. We call this 'time-delayed correlation analysis.'

In this analysis, we examine Pearson correlation coefficient, r, described as

$$r(\Delta h_{nm}, e^{t}_{ij}) = \frac{Cov(\Delta h_{nm}, e^{t}_{ij})}{SD(\Delta h_{nm}) SD(e^{t}_{ij})} \quad (1)$$

Here,

$$\Delta h_{nm} = h_n - h_m \quad (2)$$

is the difference between time-series health data h, representing the variation of the health condition, and

$$e^{t}_{ij} = e_i + e_{i-1} + ... + e_j \quad (3)$$

which is the summation of time-series lifestyle data e during a certain period, representing the accumulation of lifestyle effects. The delay is represented by retardation, $s = n - i \geq 1$ (**Figure 1**). In Eq. (1), $SD(\Delta h_{nm})$ and $SD(e^t_{ij})$ are the standard deviation of Δh_{nm} and e^t_{ij}, respectively, and $Cov(\Delta h_{nm}, e^t_{ij})$ is the covariance.

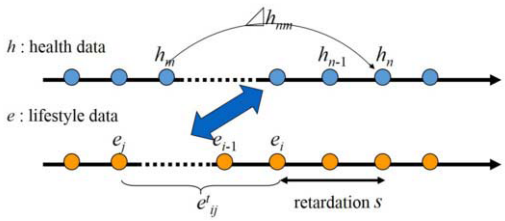


***Figure 1**: Time-delayed correlation analysis.*

Pearson correlation coefficient r is estimated for the time-series health and lifestyle data in a certain period by changing $n - m$, $i - j$, and s as parameters. If the maximum value of r is larger than a certain threshold, the lifestyle is selected as an independent variable in the data-mining process relevant to the health condition.

Rule Extraction

In the healthcare data mining process, the output (target) variable is a health condition. When the health data, h, are numeric, such as blood pressure measurements, time-series data are categorized into three classes, each representing a relative value: "higher," "moderate," or "lower." Threshold values used for this classification are determined so that the data frequency in each class is a similar number. In this study, rules are extracted with the decision tree algorithm C5.0. The study framework of time-delayed correlation analysis followed by the decision tree rule derivation is a novel approach.

Results

Time-Series Data of Blood Pressure

The time-series data of the daily systolic and diastolic home blood pressure measurements of the subject from 2004/06/01 to 2012/12/31 are shown in **Figure 2**. In addition to daily scattering, apparent seasonal variations were observed in both the systolic and diastolic blood pressure measurements. Blood pressure was higher in the winter than in the summer. Seasonal variation in the measurements has also been reported in relation to outside temperature [6].

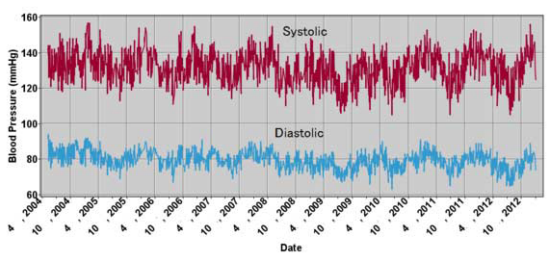


***Figure 2**: Time-series data of daily systolic and diastolic home blood pressure measurements of subject.*

To examine the seasonal variations in more detail, monthly average blood pressure values were plotted as shown in **Figure 3**.

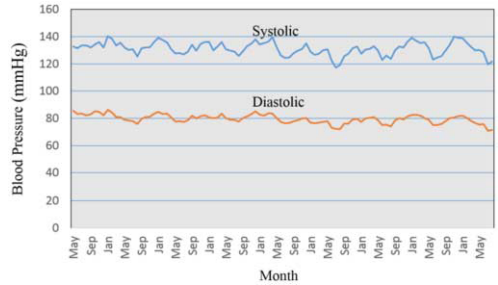


***Figure 3**: Monthly average systolic and diastolic home blood pressure measurements of subject.*

Adjustment of Seasonal Variation

We consider seasonal variation to create a bias in the correlation analysis between blood pressure and lifestyle, so we tried to adjust the blood pressure data based on the monthly average variation shown in **Figure 3**. The adjustment rate was changed monthly, as shown in Table 1.

***Table 1**: Adjustment rate of blood pressure data.*

Month	Adjustment Rate
1	0.97
2	0.98
3	0.99
4	1
5	1.01
6	1.02
7	1.03
8	1.02
9	1.01
10	1
11	0.99
12	0.98

Histograms of systolic blood pressure data before and after adjustment are shown in **Figure 4**.

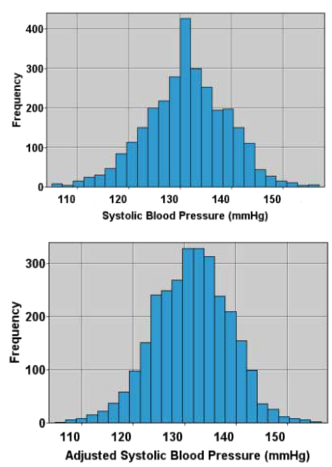


***Figure 4**: Histograms of systolic blood pressure data before and after adjustment.*

We found that the daily systolic blood pressure data approached a normal distribution when we adjusted for seasonal variations.

Histograms of diastolic blood pressure data before and after adjustment are shown in **Figure 5**. The diastolic blood pressure data also approached normal distribution after a seasonal adjustment. This implies that a seasonal adjustment is desirable to produce appropriate rules in healthcare data mining.

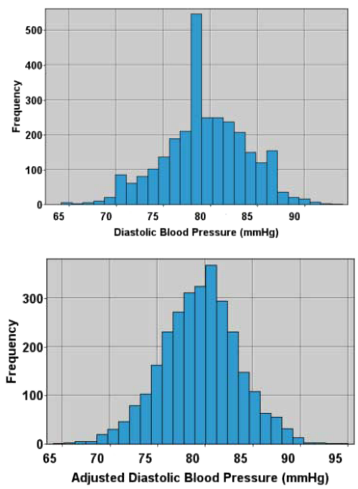


***Figure 5**: Histograms of diastolic blood pressure data before and after adjustment.*

Time-Delayed Correlation Analysis Using Adjusted Data

Time-delayed correlation analysis between blood pressure and energy expenditure/supply was performed using adjusted time-series data from 2004/06/01 to 2005/05/31 (the initial year). The parameter ranges in the analysis were $n - m = 1$–10, $i - j = 0$–9, and $s = 1$–3. Absolute values (magnitudes) of the correlation coefficients were largest when $n - m = 8$, $i - j = 1$, and $s = 1$ for the adjusted systolic blood pressure and energy expenditure/supply. Scatter plots of adjusted systolic blood pressure variation vs. energy expenditure and vs. energy supply under this condition are shown in **Figures 6** and **7**, respectively. Significant negative correlation to energy expenditure for two days and significant positive correlation to energy supply for two days were observed.

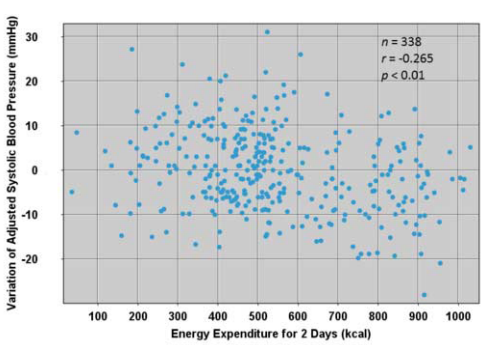


***Figure 6**: Scatter plots of adjusted systolic blood pressure variation vs. energy expenditure for two days.*

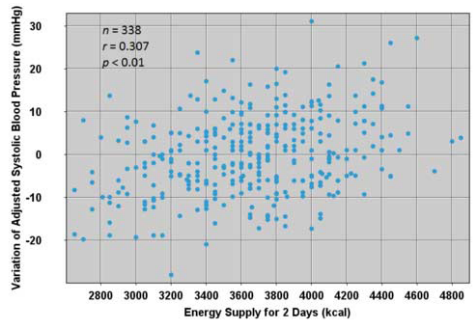


***Figure 7**: Scatter plots of adjusted systolic blood pressure variation vs. energy supply for two days.*

For comparison, the results of the time-delayed correlation analysis using non-adjusted time-series data during the same period are shown in **Figures 8** and **9**. Maximum correlation occurred when $n - m = 8$, $i - j = 1$, and $s = 1$ for both energy expenditure and supply. No significant difference was observed in the scatter plots before and after the adjustment.

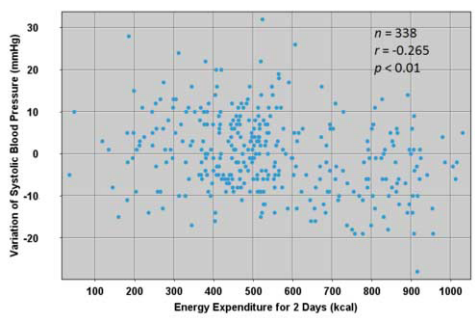


***Figure 8**: Scatter plots of systolic blood pressure variation before adjustment vs. energy expenditure for two days.*

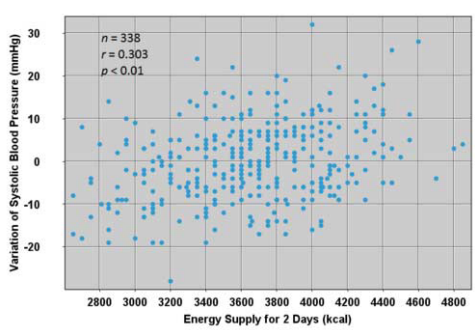


***Figure 9**: Scatter plots of systolic blood pressure variation before adjustment vs. energy supply for two days.*

This may stem from the fact that, in this analysis, periodic seasonal blood pressure variations were almost completely cancelled out by Eq. (2), since the parameter range was small ($n - m = 1$–10) compared to the period of seasonal variation. We therefore consider seasonal variation of blood pressure to have little effect on the time-delayed correlation analysis for selecting independent variables in the healthcare data mining process.

Rule Mining with C5.0 Algorithm

In the rule mining process, independent variables were energy expenditure and energy supply for two days, with a delay of one day with respect to the corresponding measurements of the blood pressure as explained in the previous section. The output (target) variable was the systolic blood pressure categorized into three classes each representing a relative value: "higher," "moderate," or "lower." Threshold values used for this classification were determined so that the data frequency in each class was similar. Rules were extracted on the basis of the time-series data from 2004/06/01 to 2005/05/31 (the initial year) using the C5.0 algorithm. An example of the decision tree for adjusted systolic blood pressure is shown in Figure 10.

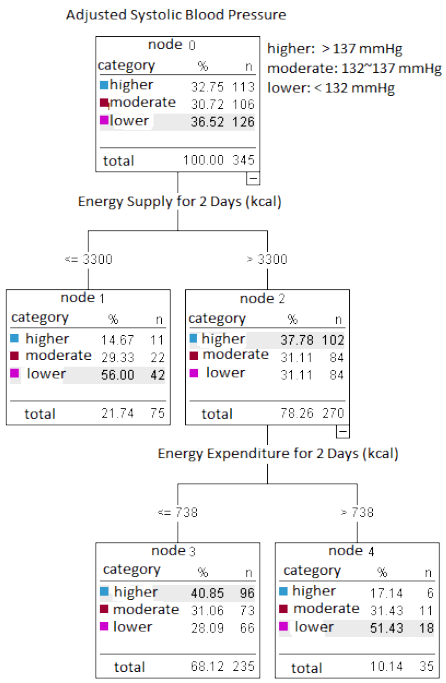


***Figure 10**: Decision tree for adjusted systolic blood pressure in relation to energy expenditure and supply.*

The decision tree produced the following rules:

1. If energy supply for two days is less than 3300 kcal, systolic blood pressure is lower (lift: 1.53).
2. If energy supply for two days is more than 3300 kcal and energy expenditure for two days is more than 738 kcal, systolic blood pressure is lower (lift: 1.41).

Here, '*lift*' is a measure of the performance of a targeting model at classifying cases. For comparison, an example of the decision tree for non-adjusted systolic blood pressure is shown in Figure 11. This decision tree produced a different rule:

3. If energy expenditure for two days is more than 570 kcal, systolic blood pressure is lower (lift: 1.43).

In general, a simpler model with fewer decision branches might be preferable. However, the rule with higher lift (1.53) involving energy supply as a lifestyle component is missing here. Thus, the adjustment of seasonal variations of blood pressure data seems to be effective at producing useful rules in healthcare data mining; although, it had little effect on the time-delayed correlation analysis to select independent variables.

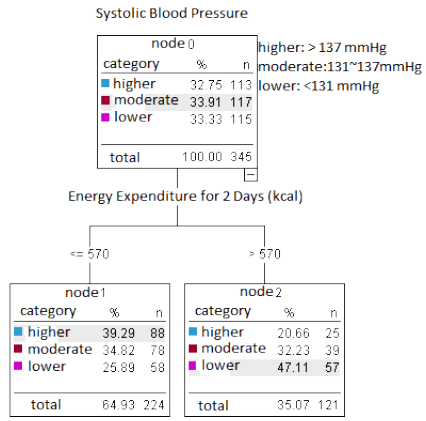


***Figure 11**: Decision tree for systolic blood pressure in relation to energy expenditure and supply.*

The adjustment of seasonal variation of blood pressure data affected the rule production. This is because the time-series blood pressure data were categorized into three classes with similar data frequency in each class for the rule mining process.

Discussion

Avoidance of Seasonal Variation Effect

Apparent seasonal variations were observed in the time-series data of an individual's home blood pressure measurements. The results of this study suggest that adjusting the seasonal variation is a desirable way to produce simpler rules between blood pressure and lifestyle (e.g. energy expenditure and supply). However, it is difficult to provide users of our PDHS with precise rules because the details of seasonal variations of personal data will only be recognized after storing the data for a long time, as shown in this study.

We tried to extract rules based on the time-series data of only three months (i.e. one season) to avoid any effect of seasonal variations. Figure 12 shows an example of the decision tree for non-adjusted systolic blood pressure based on data from 2004/06/01 to 2004/08/31 (summer in Japan). The form of the decision tree turned out to be the same as the one based on the adjusted time-series data of one year (Figure 10). Extracted rules were:

4. If energy supply for two days is less than 3700 kcal, systolic blood pressure is lower (lift: 1.29).
5. If energy supply for two days is more than 3700 kcal and energy expenditure for two days is more than 562 kcal, systolic blood pressure is lower (lift: 1.49).

These rules are similar to rules 1 and 2; although, the critical values for classification in the decision tree were somewhat different.

This result suggests that rule production based on the time-series data during only three months (one season) may be an

option for healthcare data mining. The extracted rules may still give enough information for adequate personal healthcare. At present, our PDHS produces rules once every three months for each user to avoid the effect of seasonal variations [7].

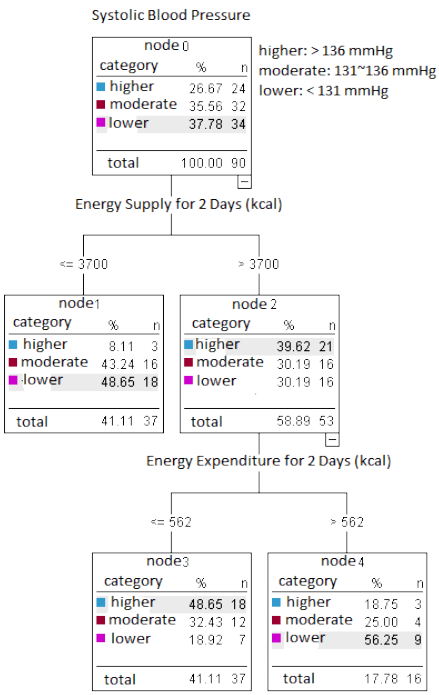


***Figure 12**: Decision tree for systolic blood pressure based on time-series data over three months.*

Threshold Values for Systolic Blood Pressure Classification

The aim of 'healthcare data mining' is to derive person-specific rules for health maintenance. In this study, the threshold values for systolic blood pressure classification were determined with a similar data frequency in each class (higher, moderate, or lower). However, the threshold values can also be determined to reflect a standard guideline in tems of defining the higher, moderate, and lower levels, unless the three classes have strongly skewed frequencies. The performance of a targeting model could be estimated by the 'lift' as well. In that case, the threshold values determined would be preferable in clinical use of our PDHS.

Summary

Long-term time-series data of an individual's home blood pressure measurements were analyzed to clarify the relation with lifestyle components, such as energy expenditure and supply. We found that the distribution of daily blood pressure data approached normal when the seasonal variation observed in the time-series data was adjusted, which demonstrates that an adjustment is desirable for producing appropriate rules between blood pressure and lifestyle components in the healthcare data mining process. The results of the healthcare data mining using the adjusted data are summarized as follows:

1. The adjustment of blood pressure data had little effect on the results of time-delayed correlation analysis to select independent variables in the data mining because, in this study, the parameter range of examining time-series correlation was small compared to the period of seasonal variation.
2. The adjustment of blood pressure data influenced the rule production process and different rules were extracted with the adjustment because the blood pressure data were categorized into three classes (higher, moderate, or lower) that were used for an output (target) variable.
3. Rules extracted after an adjustment were more useful for personal healthcare than rules extracted before the adjustment.
4. Rule production based on the time-series data of just three months (one season) is an option for healthcare data mining to avoid the effects of seasonal variation.

Acknowledgements

This work was supported in part by Grants-in-Aid for Science Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

[1] Kumpusch H., Hayn D., Kreiner K., Falgenhauer M., Mor J, and Schreier G. A mobile phone based telemonitoring concept for the simultaneous acquisition of biosignals and physiological parameters. Proc. 13th World Congress on Medical and Health Informatics, 2010; pp. 1344–48.

[2] Zapata B.C., Ninirola A.H., Fernandez-Aleman J.L., and Toval A. Assessing the privacy policies in mobile personal health records. Proc. 36th Annual International Conference of the IEEE EMBS, 2014; pp. 4956–59.

[3] Kantoch E., Augustyniak P., Markiewicz M., and Prusak D. Monitoring activities of daily living based on a wearable wireless body sensor network. Proc. 36th Annual International Conference of the IEEE EMBS, 2014; pp. 586–89.

[4] Takeuchi H., Mayuzumi Y., Kodama N., and Sato K. Personal healthcare system using cloud computing. Proc. 14th World Congress on Medical and Health Informatics, 2013; p. 936.

[5] Takeuchi H., Kodama N., Hashiguchi T., and Hayashi D. Automated healthcare data mining based on a personal dynamic healthcare system. Proc. 28th Annual International Conference of the IEEE EMBS, 2006; pp. 3604–07.

[6] Hozawa A., Kuriyama S., Shimazu T., Ohmori-Matsuda K., and Tsuji I. Seasonal variation in home blood pressure measurements and relation to outside temperature in Japan, Clin. Exp. Hypertens. 33(3), 2011; pp. 153–58.

[7] Takeuchi H. and Kodama N. Validity of association rules extracted by healthcare-data-mining. Proc. 36th Annual International Conference of the IEEE EMBS, 2014; pp. 4960–63.

Correspondence

Hiroshi Takeuchi is with the Department of Healthcare Informatics, Takasaki University of Health and Welfare, 37-1, Nakaorui-machi, Takasaki-shi, Gunma, 370–0033, Japan.

Phone: +81-27-352-1290
FAX: +81-27-353-2055
E-mail: htakeuchi@takasaki-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-756

The Improvement of Dental Posture Using Personalized Biofeedback

Bhornsawan Thanathornwong[a], Siriwan Suebnukarn[b]

[a] *Faculty of Dentistry, Srinakarinwirot University, Bangkok, Thailand*
[b] *Faculty of Dentistry, Thammasat University, Pathumthani, Thailand*

Abstract

Background: Dentists are subject to staying in static or awkward postures for long periods due to their highly concentrated work. Objectives: This study describes a real-time personalized biofeedback system developed for dental posture training with the use of vibrotactile biofeedback. Methods: The real-time personalized biofeedback system was an integrated solution that comprised of two components: 1) a wearable device that contained an accelerometer sensor for measuring the tilt angle of the body (input) and provided real-time vibrotactile biofeedback (output); and 2) software for data capturing, processing, and personalized biofeedback generation. The implementation of real-time personalized vibrotactile feedback was computed using Hidden Markov Models (HMMs). For the test case, we calculated the probability and log-likelihood of the test movements under the Work related Musculoskeletal Disorders (WMSD) and non-WMSD HMMs. The vibrotactile biofeedback was provided to the user via a wearable device for a WMSD-predicted case. In the system evaluation, a randomized crossover trial was conducted to compare dental posture measure using tilt angles of the upper back and muscle activities of those dental students that received vibrotactile biofeedback from the system with the control group against the dental students who received no feedback. Results: The participants who received feedback from the system had a lower tilt angle at 10^{th}, 50^{th}, and 90^{th} percentiles of $Back_x$ and $Back_y$, as well as muscular load, which were statistically different ($p < 0.05$) from those who received no feedback from the system. Conclusions: The results presented here demonstrate that a personalized biofeedback system for posture training in dental students is feasible and associated with quantitative improvements of the dental posture.

Keywords:

Hidden Markov Model; Musculoskeletal disorders; Vibrotactile biofeedback; Personalization.

Introduction

A large number of dentists suffer from musculoskeletal problems later in their professional lives. Some dentists have milder forms of musculoskeletal problems, while others have much more severe forms. A proportional correlation between the number of disorders and the years of clinical experience have been reported.[1] Dentists are subject to staying in static or awkward postures for long periods due to their highly concentrated work and the restriction of the oral cavity. According to Rising et al.,[2] more than 70 percent of dental students reported neck, shoulder, and lower back pain by their third year of dental school. In order to prevent this; correct dentist posture must be established early among dental students. Therefore, the correct posture must be stressed in dental schools. Although most schools teach the correct and ideal dentist posture and positions, they are not always applied by the dental students themselves.

A complex interplay of feedback and feed-forward control ensures the natural ability of the human body to maintain an upright stance, and to stabilize during movement.[3] One approach to improving balance which has been widely used in physical therapy and rehabilitation involves feeding back to the central nervous system supplementary environmental information about body motion. This supplemental information may come from a therapist, laboratory equipment, or artificial sensors.[4] Biofeedback (BF) systems for postural control are aimed at providing additional sensory information to supplement natural sensory information and improve human balance. Since the experimentation of biofeedback systems for postural control began, tactile and audio biofeedback have received much less attention than visual biofeedback. Nevertheless, in the last few years, interest in tactile- and audio-biofeedback systems for postural control has been renewed,[5,6] partially due to advances in technology for real-time processing and movement sensing, and to new trends in wireless wearable devices that can be worn during daily activities. Audio-biofeedback experimentation was carried out by Chiariet et al. [6] and Hegeman et al., [7] who developed audio BF devices able to encode trunk movement information into a sound. In 2001, Wall et al. [5] developed a device able to provide tactile feedback of trunk tilt by vibrating tactors that the subjects wore around their trunk. This study showed how vibrotactile feedback might improve balance performance in healthy subjects, as well as the possible usefulness and validity of this system as a prosthesis for people with pathologies that impair balance.

The use of biofeedback has been offered in the past as an instrument for training that enables an individual to learn how to change physiological activity or behavior for the purposes of improving performance. In therapeutic applications, biofeedback training of balance and posture has shown to be effective for posture control in adolescent scoliosis,[8] and has also decreased the fall rate in elderly patients with peripheral neuropathy.[9] In patients with bilateral vestibular loss,[10] biofeedback training was also found useful in enhancing postural stability even under challenging standing conditions, beyond the effect of practice alone.[10-12]

The aims of this study were to investigate the manner and tasks in which the personalized biofeedback system can be used to enhance postural control in dental operation, to explore the feasibility of using a personalized biofeedback system for posture training of dental students, and to preliminarily assess the usability and efficacy of a personalized biofeedback paradigm on a group of dental students. This study describes a real-time personalized biofeedback system developed for posture training with the use of vibrotactile biofeedback. The value of this study is that

no previous studies have shown how a personalized biofeedback system could be employed by dental students for monitoring their posture, as well as helping them to self-correct their posture and movements in order to minimize the risk of acquiring musculoskeletal problems. Although our focus of attention is in the area of dentistry, the proposed system could be applied to other domains that require self-monitoring and correction of posture.

Methods

Real-time personalized biofeedback system

The real-time personalized biofeedback system was an integrated solution comprised of two components (Figure 1): 1) A wearable device that contained an accelerometer sensor for measuring the tilt angle of the body (input) and provided real-time vibrotactile biofeedback (output). The angles to be measured included: flexion and extension, left and right lateral flexion during the operation. The system took those angles as input to the software. 2) The software was used for tilt angle data capturing, data processing, and personalized biofeedback generation. The implementation of real-time personalized vibrotactile feedback was computed using Hidden Markov Models (HMMs).[13] HMMs were used to predict whether the dental student was likely to acquire Work-related Musculoskeletal Disorders (WMSD) by comparing the dental student's movement patterns with WMSD and non-WMSD HMMs learned from previous data. The vibrotactile biofeedback was provided to the user via a wearable device for a WMSD-predicted case. The idea was to encourage dental students to correct upper back movement themselves after receiving vibrotactile biofeedback during their dental work.

Device

The wearable device developed by our group consisted of an electronic system that produced a voltage signal in reaction to body movements. It consisted of an ADXL345 3-axis accelerometer with high-resolution (13-bit) measurement at up to ±16g and a 12.5-400Hz bandwidth response. An accelerometer sensor was used for tracking body tilt angles. The accelerometer sensor presented two reading outputs, one for the X_{out} and another for Y_{out}, and a power supply voltage input of 2.0-3.6V. The expected values for X_{out} and Y_{out} were in the digital IO voltage range of 1.8-2.5V. The sensor consisted of a structure with a capacitive sensing cell (g-cell) and signal conditioning to detect small displacements. The signals from the accelerometer sensors were amplified and converted into digital signals through a data acquisition card (13-bit resolution) connected to a computer (Laptop computer with a Core i5 processor, running Microsoft® Windows 7). The accelerometer mounted on a circuit card was used as an inclinometer to calculate body tilt angles during the dental operation (Figure 2).

Software

The software was developed using Visual Studio to control and process the arrival of the signals obtained through a flash memory to the computer. The software configured the entrance channels and was programmed considering the pins where the sensors had been connected to the data acquisition card through an analogical signal interface cable. As soon as the signals arrived at the acquisition card, they were available in their respective channels. The voltages were used with calibration values in order to obtain the values of flexion/extension of the body tilt in degrees. Results from the data processing were stored in a database.

Figure 1 - Personalized biofeedback system

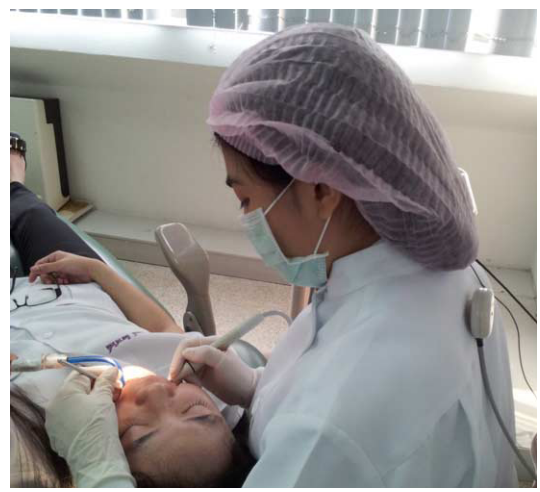


Figure 2 - The wearable device containing an accelerometer sensor and vibrotactile biofeedback module was attached to the operator's gown to measure the tilt angles of the upper back. The device was lightweight and did not interfere with the dental operation.

Personalized biofeedback

We propose a HMM as a statistical tool to objectively assess dental posture based on the measured data about the operator's tilt angles. HMMs have been used extensively, and have shown to be effective in applications such as gesture recognition [14] and speech recognition.[13] They also have been used for modeling human operator skills and transferring them to robots.[15] Recently, HMMs have been applied to model complex tasks such as surgery (specifically in automatic assessment of surgical performance in laparoscopy),[16] pelvic examination,[17] and mastoidectomy.[18] These applications suggest that HMMs have high potential to provide accurate models for assessing dental posture.

We conducted an experiment to test the ability of a machine learning technique, the HMM, to recognize and classify an observed dentist movement patterns as WMSD or non-WMSD, based on a set of recorded important features. The training data were obtained from fifty general dentists. Thirty

dentists were identified as WMSD, and 20 as non-WMSD according to Kroemer's guidelines.[19] Kroemer's guidelines classified WMSD into 3 stages: Stage 1 is characterized by local aches and tiredness during the working hour, which usually abate overnight and with days away from work; Stage 2 has symptoms of tenderness, swelling, numbness and pain that starts early in the work shift and does not abate overnight; Stage 3 is characterized by symptoms that persist at rest and during the night. In our study Stage 2 and 3 were considered to be in the WMSD group and Stage 1 or no symptoms were considered to be in the non-WMSD group. Informed consent was provided upon recruitment to the study. The operation selected for our study was scaling on the upper right quadrant of mild gingivitis patients during their routine schedule. During the work, we continuously stored the data on the right and left lateral flexion, and flexion and extension of the upper back ($Back_x$ and $Back_y$).

Once the models were trained, we calculated the probability and log-likelihood of the test movements under the WMSD and non-WMSD HMMs using the forward algorithm as described previously to find the model that best describes the test movement data. If the log-likelihood of the test movements under WMSD HMM was greater than under the non-WMSD HMMs, the system classified the test movement as WMSD; otherwise, it was classified as non-WMSD. The personalized vibrotactile biofeedback was provided to the user via a wearable device for a WMSD predicted case.

We used discrete HMMs to model the system. To validate the model precision, we performed five-fold cross validation. A different *k*-means clustering algorithm was used for every cross validation fold and the same *k*-means for the WMSD and non-WMSD model in the same fold. For each fold, we trained the WMSD HMM with four WMSD and four non-WMSD sequences. To determine the accuracy of the method, after training the two HMMs in each fold, we fed the test WMSD and non-WMSD data to each model. The average log likelihood of all sequences across all five folds for the two HMMs is shown in Table 1. In every cross validation fold, the log likelihood of every test sequence under its corresponding HMM was higher than that under the other HMM. These results demonstrate the ability of the HMM to distinguish between WMSD and non-WMSD movement with 100% accuracy.

Table 1 - Average log likelihood results for WMSD and non-WMSD movement sequences

	Log likelihood for WMSD HMM	Log likelihood for non-WMSD HMM
WMSD	-3.475×10^3	-2.121×10^6
Non-WMSD	-6.142×10^5	-3.584×10^3

System evaluation

Participants and design

In this study, a randomized crossover trial was conducted to compare tilt angles of the upper back and muscle activities of the participants that received vibrotactile biofeedback from the system with the control group who received no feedback. Upper back tilt angles while performing scaling on upper 1st and 2nd molars were measured. We hypothesized that the participants that received vibrotactile biofeedback from the system will improve their postures (decreased $Back_x$ degree and $Back_y$ degree, and upper trapezius muscle activity). We recruited sixteen dental students (8 females and 8 males) aged between 21 and 23. The choice of at least 16 participants per group was based on a 2-tailed test, with $\alpha = 0.05$ and power $(1-\beta) = 0.80$. The inclusion criteria were that the participants performed a minimum of 6 h of dental practice a day. They were not admitted to the study if they received below 70 percent marks in knowledge assessment of dental ergonomics. Participants answered a questionnaire about their health and workplace. None of them were excluded from the group on health grounds. The study was approved by the institutional Ethical Review Board. A written consent form was provided by all participants. Participants were randomly assigned into a 2x2 crossover trial using a computer–generated randomization schedule to each of two sequences of working. Participants in the experiment group received vibrotactile biofeedback from the system after finishing scaling (Feedback), while those who were in the control group received no feedback (Control).

Data analysis

The primary outcome measures were the mean values of $Back_x$ degree and $Back_y$ degree. The secondary outcome measure was upper trapezius muscle activity measured using electromyography (EMG). The primary and secondary outcome measures were recorded two times: after finishing scaling on the 1st molar (Pre-test) and finishing scaling on the 2nd molar (Post-test). The primary outcome measures were the angle in degree at 10th, 50th, and 90th percentile of $Back_x$ and $Back_y$ accurate to 0.01 degree. The secondary outcome measures were raw EMG signals processed with BioPAK™ Program software V 7.2 (BioResearch Associates Inc., RI, USA). Raw EMG data was transformed into frequency domain and band-pass filtered at a high-pass frequency of 10 Hz and a low-pass frequency of 500 Hz. Then, the data was inverse-transformed to time domain for further analysis. Accordingly, filtered signals were full-wave rectified and averaged across data collection period.Descriptive statistics were used to evaluate the effects of real-time personalized biofeedback system. Average, standard deviations and percentile were extracted as well as the percent change of post-test from pre-test. Personalized biofeedback effects (control vs. feedback group) were evaluated using the Paired t-test and were assumed to be significant at $p < 0.05$ (two-side). All analyses were conducted with the statistical package for the Social Science version 21.0 (SPSS, Chicago, IL).

Results

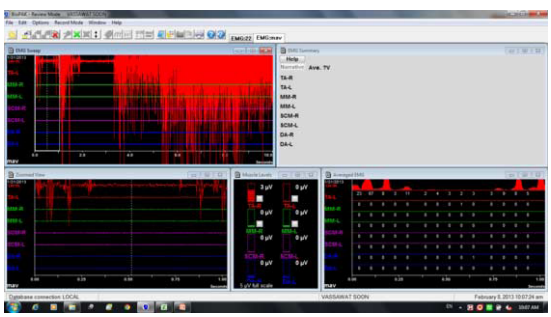
Dental students that received feedback from the system had lower tilt angles at 10th, 50th, 90th percentile of $Back_x$ and $Back_y$, as well as muscular load which were statistically different ($p < 0.05$) from those who received no feedback from the system. A significant improvement of 3.62-8.47 degrees was seen for forward movements, and 6.12-8.88 degrees for sideways movements in the group that received feedback. A sample EMG that shows muscular load for the right trapezius from one participant doing dental procedure is shown in Figure 3. There was lower muscular load for the right trapezius while receiving vibrotactile biofeedback from the system (Figure 3b), as compared to receiving no feedback (Figure 3a).

Discussion

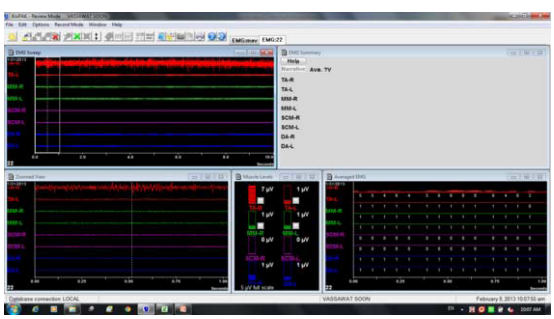
In a previous study, a real-time system with assistive feedback for postural control in rehabilitation found that the system was suitable for clinical applications pertaining to postural control improvements.[20] The limitations were that custom-developed software might not effectively be applied to all patients. Another study applied a wearable real-time

intelligent posture corrective system using vibrotactile feedback to improve ankle proprioception in wobble board training. A fuzzy inference system was used to determine the quality of postural control, based on inertial measurement units-acquired measurements of trunk and wobble board. The results observed an improvement in postural control with biofeedback intervention, demonstrating the success of the prototype built for improving postural control in rehabilitative and preventive applications.[21]

In our study besides posture position, the muscle activity was also assessed using EMG. During dental work there was higher load for the right trapezius in the control group, as compared to the feedback group. Our results are in line with the study on quantifying work load in neck, shoulders and wrists in female dentists.[22] It was found that dentists were exposed to high load on the trapezius muscles bilaterally, and steep, prolonged forward bending of the head.



(a)



(b)

Figure 3 - Relationships between muscular activity and maximal voluntary contraction (MVC) for right trapezius muscle of (a) control and (b) feedback from one participant. The upper-left window shows EMG sweep in 10 sec. of right trapezius muscle (read from TA-R electrode). The lower-left window is the zoomed view of EMG in 1 sec. The lower-middle window shows a single muscular activity level selected from the zoomed view. The lower-right window shows average EMG in 1 sec.

As hypothesized, the postural training using personalized biofeedback reduced muscle activity significantly. Trapezius muscle activity has been studied previously in dental work research due to the discomfort that is experienced in the neck/shoulder region.[22,23] The magnitude of EMG signals of the Trapezius muscles while working on the dental procedure were compared between sitting in the conventional chair and General Chair (Ergonomically Designed Chair EDC;) designed by strong support over the arms and trunk.[24] The results showed that the chairs are designed specifically to reduce activity of the muscles significantly and MVE% is close to the average value that is derived from data received participants' using personalized biofeedback in this study.

A study on biofeedback with muscle activity by Palmerud et al. [25] examined with electromyography in abducted arm positions. By using feedback techniques, they found that the subjects could reduce EMG activity voluntarily by 56% in the trapezius muscle while keeping different static postures. When compared with this study, the values of the EMG activity from a feedback group decrease by 53.15%. Thus, while you are working in the appropriate position, the function of the muscle decreases.

Many research experiments have been introduced using biofeedback information. Biofeedback systems have become a prominent component in motor training and rehabilitation. Alahakone et al. [20] found that instantaneous feedback provided via vibration stimulus can reduce postural sway based on trunk tilt measurements. Hence, the system's pertinence to comparable approaches employed in sports training and rehabilitation is apparent. Correspondingly, vibrotactile biofeedback improves gait in patients with unilateral vestibular loss. Results showed an immediate improvement in postural stability (reduction of lateral center of mass displacement, trunk tilt and medial-lateral step width), that was significantly larger than effects of practice alone.[26] In addition to evaluate the effectiveness of an audiobiofeedback system for improving balance in patients with bilateral vestibular loss. Audiobiofeedback improved stance stability of participants with bilateral vestibular loss by increasing the amount of postural corrections.[27]

To our knowledge, this is the first intervention trial using a personalized biofeedback system for training posture in dental students. The present study aimed to explore if this training method is feasible for dental students. Future studies should include a larger sample of dental students and dentists, as well as other occupations (e.g., surgeons, computer users, or drivers). Training with the personalized biofeedback system teaches participants new strategies of movement that could be applied to real-life situations. In this sense, the personalized biofeedback system may have an advantage over other technologies used by dentists, by enhancing motor learning through feedback on knowledge of performance, and knowledge of results. Further studies are needed to look at the possibility of using a personalized biofeedback system for daily training.

Conclusions

The results presented here demonstrate that the personalized biofeedback system for posture training in dental students is feasible, and associated with quantitative improvements of dental posture. This may be viewed as a promising first step to implement posture training strategies to minimize work-related musculoskeletal disorders in dentists and other healthcare workers.

Acknowledgements

This study was supported by Thailand Office of Higher Education Commission, National Research Council of Thailand, and the Center of Excellence in Biomedical Engineering of Thammasat University, Thailand.

References

[1] Szymanska J. Disorders of the musculoskeletal system among dentists from the aspect of ergonomics and prophylaxis. Ann Agric Environ Med. 2002;9:169-73.

[2] Rising DW, Bennett BC, Hursh K, Plesh O. Reports of body pain in a dental student population. J Am Dent Assoc. 2005;136:81-6.

[3] Giansantia D, Dozzab M, Chiarib L, Maccionia G, Cappello A. Energetic assessment of trunk postural modifications induced by a wearable audio-biofeedback system. Medical Engineering & Physics. 2009;31:48–54.

[4] Wall III C, Weinberg MS. Balance prostheses for postural control. IEEE Eng Med Biol Mag. 2003;22(2):84–90.

[5] Wall III C, Weinberg MS, Schmidt PB, Krebs DE. Balance prosthesis based on micromechanical sensors using vibrotactile feedback of tilt. IEEE Trans Biomed Eng. 2001;48(10):1153–61.

[6] Chiari L, Dozza M, Cappello A, Horak FB, Macellari V, Giansanti D. Audio-biofeedback for balance improvement: an accelerometry-based system. IEEE Trans Biomed Eng. 2005;52(12):2108–11.

[7] Hegeman J, Honegger F, Kupper M, Allum JH. The balance control of bilateral peripheral vestibular loss subjects and its improvement with auditory prosthetic feedback. J Vestib Res. 2005;15(2):109–17.

[8] Wong MS, Mak AF, Luk KD, Evans JH, Brown B. Effectiveness of audio-biofeedback in postural training for adolescent idiopathic scoliosis patients. *Prosthet Orthot Int*. 2001;25:60-70.

[9] Wu G. Real-time feedback of body centre of gravity for postural training of elderly patients with peripheral neuropathy. *IEEE Trans Rehabil Eng*. 1997;5:399-402.

[10]Dozza M, Chiari L, Horak FB. Audio-biofeedback improves balance in patients with bilateral vestibular loss. *Arch Phys Med Rehabil.* 2005;86:1401-03.

[11]Dozza M, Chiari L, Chan B, Rocchi L, Horak FB, Cappello. A**:** Influence of a portable audio-biofeedback device on structural properties of postural sway. *J Neuroeng Rehabil.* 2005;2:13.

[12]Horak FB, Dozza M, Peterka R, Chiari L, Wall C. Vibrotactile biofeedback improves tandem gait in patients with unilateral vestibular loss. *Ann N Y Acad Sci.* 2009;1164:279-81.

[13]Rabiner LR. A tutorial on hidden Markov models and selected application in speech recognition. In: Waibel A, Lee KF, editors. Proceedings of the IEEE. San Mateo: Kaufmann; 1989. pp. 257–86.

[14]Rhienmoraa P, Haddawyb P, Suebnukarnc S, Dailey MN. Intelligent dental training simulator with objective skill assessment and feedback. Artificial Intelligence in Medicine. 2011;52(2): 115–21.

[15]Hovland GE, Sikka P, McCarragher BJ. Skill acquisition from human demonstration using a hidden Markov model N. Caplan, C.G. Lee (Eds.), Proceeding of the IEEE international conference on robotics and automation (ICRA-96), vol. 3, IEEE, Piscataway, NJ (1996), pp. 2706–2711.

[16]Rosen J, Solazzo M, Hannaford B, Sinanan M. Objective laparoscopic skills assessments of surgical residents using hidden Markov models based on haptic information and tool/tissue interactions. Studies in Health Technology and Informatics, Medicine Meets Virtual Reality, Newport Beach, CA. 2001;81:417–23.

[17]Mackel T, Rosen J, Pugh C. Data mining of the e-pelvis simulator database: a quest for a generalized algorithm for objectively assessing medical skill Studies in Health Technology and Informatics. 2006;119:355–60.

[18]Sewell C, Morris D, Blevins NH, Dutta S, Agrawal S, Barbagli F et al. Providing metrics and performance feedback in a surgical simulator. Computer Aided Surgery. 2008;13(2):63–81.

[19]Kroemer KHE. Job design. In: Kroemer KHE, editor. Ergonomic design of material handling systems. 1st ed. New York: Lewis publishers; 1997. p. 360-435.

[20]Alahakone AU, Senanayake SMNA. A Real-Time System With Assistive Feedback for Postural Control in Rehabilitation. Mechatronics. 2010;15:226-33.

[21]Gopalai AA, Senanayake SMNA. A Wearable Real-Time Intelligent Posture Corrective System Using Vibrotactile Feedback. Mechatronics. 2011;16:827-34.

[22]Akesson I, Hansson GA, Balogh I, Moritz U, Skerfving S. Quantifying work load in neck, shoulders and wrists in female dentists. Int Arch Occup Environ Health. 1997;69:461-74.

[23]Finsen L, Christensen H. A biomechanical study of occupational loads in the shoulder and elbow in dentistry. Clin Biomechanics. 1998;13:272-9.

[24]Haddad O, Sanjari M, Amirfazli A, Narimani R, Parnianpour M. Trapezius Muscle Activity in using Ordinary and Ergonomically Designed Dentistry Chairs. Int J Occup Environ Med. 2012;3(2):76-83.

[25]Palmerud G, Kadefors R, Sporrong H, Ja¨rvholm U, Herberts P. Voluntary redistribution of muscle activity in human shoulder muscles. Ergonomics. 1995;38:806-15.

[26]Horak FB, Dozza M, Peterka R, Chiari L, and Wall C. Vibrotactile Biofeedback Improves Tandem Gait in Patients with Unilateral Vestibular Loss. Ann N Y Acad Sci. 2009;1164:279-81.

[27]Dozza M, Chiari L, Horak FB. Audio-biofeedback improves balance in patients with bilateral vestibular loss. Arch Phys Med Rehabil. 2005;86:1401-403.

Address for correspondence

Siriwan Suebnukarn, DDS, PhD

Faculty of Dentistry, Thammasat University

Pathumthani, Thailand, 12121

Tel: +66-1-6425582 Fax: +66-2-9869205

E-mail: ssiriwan@tu.ac.th

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-761

Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter

Xujuan Zhou[a], Enrico Coiera[a], Guy Tsafnat[a], Diana Arachi[a], Mei-Sing Ong[a,b], Adam G. Dunn[a]

[a] *Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University NSW, Australia*
[b] *Children's Hospital Informatics Program, Boston Children's Hospital, Boston MA, United States*

Abstract

The manner in which people preferentially interact with others like themselves suggests that information about social connections may be useful in the surveillance of opinions for public health purposes. We examined if social connection information from tweets about human papillomavirus (HPV) vaccines could be used to train classifiers that identify anti-vaccine opinions. From 42,533 tweets posted between October 2013 and March 2014, 2,098 were sampled at random and two investigators independently identified anti-vaccine opinions. Machine learning methods were used to train classifiers using the first three months of data, including content (8,261 text fragments) and social connections (10,758 relationships). Connection-based classifiers performed similarly to content-based classifiers on the first three months of training data, and performed more consistently than content-based classifiers on test data from the subsequent three months. The most accurate classifier achieved an accuracy of 88.6% on the test data set, and used only social connection features. Information about how people are connected, rather than what they write, may be useful for improving public health surveillance methods on Twitter.

Keywords:
Machine learning; Social media; HPV vaccines; Public health surveillance; Twitter messaging; Text mining.

Introduction

Social media surveillance applications that provide value to public health include surveying demographics, estimating population-wide sentiment about public health issues like vaccines, forecasting influenza outbreaks, and producing spatial indicators of language, behaviour, and mood [1-7]. One of the specific problems associated with using Twitter for online surveillance is the brevity and non-standard text structures of Twitter posts (tweets), which limit the text fragments that can be used to train classifiers, and may limit performance [8, 9].

We hypothesized that connections between users on social media may help to improve surveillance methods for the following reasons: (a) homophily – where people tend to form connections with others who share similar attributes or opinions [10-12]; (b) contagion of opinions – where social connections represent the conduits through which information flows, influencing and shaping opinions [13-15]; and (c) temporal dynamics – where user relationships may change more slowly than the content in the topics being discussed.

To test the hypothesis that social connections could improve the performance of opinion classification methods, we considered a classification task in the surveillance of anti-vaccine rhetoric about human papillomavirus (HPV) vaccines on Twitter. The growth of anti-vaccine rhetoric in the media is an international problem [16, 17]. HPV vaccines are a relatively recent introduction to the armament of public health, and uptake is highly variable by country, demographic, and location [18]. Anti-vaccine rhetoric specifically directed at HPV vaccines is present in media articles and websites [19-21], and this appears to have the capacity to alter vaccine acceptance and decision-making [22].

The aim of this study was to determine if information about social connections could be used to improve the performance of classifiers intended for ongoing use in public health surveillance, using anti-vaccine rhetoric as an example.

Methods

Study Data

English-language tweets (42,533 tweets) containing keywords related to HPV vaccines were collected between October 1, 2013 and March 31, 2014. We identified these tweets by searching for combinations of keywords (HPV, vaccine, Gardasil, Cervarix, vaccination, cervical, cancer) via repeated calls to the Twitter application programming interface (API), in accordance with the terms of service for Twitter developers. For each of the users responsible for the tweets (21,166 users), the sets of users they followed (*sources*), as well as the sets of users that followed them (*followers*), were accessed through separate API requests and recorded soon after the first time they tweeted about HPV vaccines in the six-month period.

We split the six months of data into two distinct but contiguous three-month periods and randomly sampled tweets for use in the classifier training (1050 tweets from October 2013 to December 2013) and testing (1100 tweets from January 2014 to March 2014). Two investigators (DA and AD) independently rated each tweet as having presented an anti-vaccine opinion or otherwise. Agreement between the investigators was 95% in the training period (Cohen's kappa 0.88; $p<0.001$), and 95% in the testing period (Cohen's kappa 0.86; $p<0.001$). Disagreements were resolved by discussion. Tweets were removed if they were deleted or the user had become protected or suspended, or if the text was identical after pre-processing. Final samples used included 884 and 907 tweets in the training and testing period, of which 247 (28%) and 201 (22%) were labelled as anti-vaccine, respectively.

Data pre-processing

We pre-processed the text (content) to remove punctuation, words that were unlikely to confer meaning (e.g. 'and'), and other non-word elements (e.g. URLs). The remaining text was used to produce sets of unigrams (one word) and bigrams (two contiguous words), which were then available for use in the classifier training. An example is given in Figure 1.

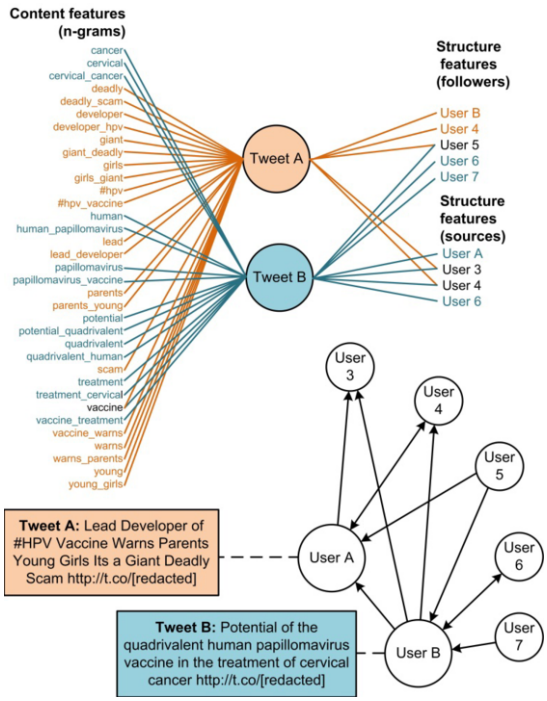


Figure 1– The text is decomposed into n-gram features (left). The follower network for the two users posting the tweets is decomposed into source and follower features (right)

We then determined the source and follower relationships among the set of 21,166 users who tweeted at least once about HPV vaccines. Social connection features were constructed directly from the follower relationships between users. An example of the decomposition is given in Figure 1.

Statistical analysis of content and connection features

Using the sample from the training period, we identified content and connection features that were significantly over-represented in one of the two classes by applying Fisher's exact tests to each feature and then a Bonferroni correction. To examine how low-frequency features might affect the performance of classifiers in the training and testing periods, we also relaxed the inclusion criteria to create alternative sets of features to be used as inputs to the classifier training. The first included all features for which the p-values were less than 0.05, and the second included the set of all features represented in at least three tweets in the sample.

The frequencies of the features exhibiting significant associations in period one were then compared to their frequencies in period two, to measure how the associations may degrade over time. For each of the sets of significant features, we calculated the proportion of features that retained a significant over-representation in the same class.

Classification algorithms

To demonstrate how the temporal variation in the content and connection information might affect the performance of supervised machine learning classifiers, we demonstrated the approach by constructing classifiers using support vector machine (SVM) methods. SVM classifiers have been widely applied in text-based classification [23, 24], and sentiment analysis [25, 26], and are considered the standard and the most appropriate classifier for unbalanced datasets and a large number of features. We chose to apply radial basis function kernels [27], and used consistent parameter values across all the classifiers in order to avoid retrospectively influencing the reporting of the performance.

Feature selection methods are heuristics that are used in the training of machine learning classifier to improve performance by ignoring features that are not useful, and including combinations of features that are best able to discriminate between classes. We applied a hybrid method of forward selection and backward elimination [28, 29].

Classifier construction and testing in the training period was undertaken using a ten-fold cross validation. Note that we determined the potential features using the statistical analysis covering the entire training period. In the testing period (the subsequent three months), the classifiers were applied directly to the full set of tweets from the period as a holdout set, in order to examine the temporal degradation. To compare the classifier performance from training to testing periods, we calculated the standard performance measures: precision, recall, accuracy, and F_1-score. The research project was approved by the Human Ethics Advisory Panels of The University of New South Wales and Macquarie University.

Results

Temporal degradation in content and connection features

From the set of 42,533 unique English-language tweets spanning six months, a random sample of 2150 were extracted and then manually labelled as anti-vaccine or otherwise. After pre-processing, 884 tweets remained in the sample in the first six months (training period), and these came from 877 unique users. Applying Fisher's exact tests and Bonferroni corrections, we identified text fragments and social connections that were found to be significantly more frequent in one class relative to the other (Table 1).

Table 1– The frequency of content and connection features compared across the two periods

Characteristic Set	Number of unique tweets	Number of anti-vaccine tweets (%)	Number of significant features	Features that were no longer significant (%)	Features that switched direction (%)	Features that were still significant (%)
Bigrams (content)						
Bonferroni-corrected	884	247 (28%)	25	24 (96%)	0	1 (4%)
p-value <0.05	884	247 (28%)	232	228 (98%)	2 (0.86%)	2 (0.86%)
Followers (connections)						
Bonferroni-corrected	877	220 (25%)	73	0	0	73 (100%)
p-value <0.05	877	220 (25%)	542	220 (41%)	0	322 (59%)
Sources (connections)						
Bonferroni-corrected	877	220 (25%)	82	2 (2.5%)	0	80 (98%)
p-value <0.05	877	220 (25%)	494	183 (37%)	0	311 (63%)

When the same features were then compared across classes in the tweets from the testing period, the comparison showed that only 1 of 24 (4%) of the content features were also significant in the subsequent three months (Figure 2). In comparison, 80 of 82 (98%) of the connection-based source features were also significant in the subsequent three months, as well as 73 of the 73 (100%) of the connection-based follower features (Table 1). The results show that very few text-based features retained their significant differences in the testing period, while social connections nearly always retained their significant differences in the testing period.

It might be expected that the reason why connection features are stable from one period to the next is because the same users are responsible for anti-vaccine tweets in both periods. However, among the users in the tweets sampled from the training period (877 users), and the testing period (797 users), only 4.1% of the users (66 of 1,608) appear in both samples. Extending this analysis to consider all original tweets in the two periods and not just the sampled sets, only 11.3% of users (2,382 of 21,166) posted tweets about HPV vaccines in both periods. The small overlap suggests that the connection features were stable across the two periods not as a consequence of tweets being posted by the same users, but because users posting about HPV vaccines for the first time in the six month period often followed the same accounts as other users who expressed similar opinions.

Classifier training and testing in period one

The classifiers trained using only connection features produced similar levels of accuracy (often with higher precision and lower recall) to the classifiers that were trained using only content features (Table 2). The best-performing classifier that only used connection features achieved an accuracy of 89.4% (95% CI 87.4-91.4), which was roughly equivalent to the best-performing classifier trained using only content features (89.8% accuracy; 95% CI 87.9-91.8). The overall best-performing classifier in the training period was constructed from both content and connection features (94.4% accuracy; 95% CI 93.1-96.3), and used 23 social connections and 28 text-based features.

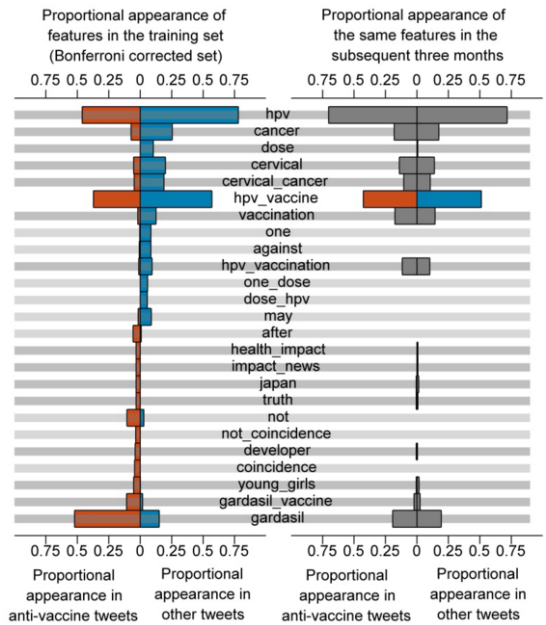


Figure 2– The proportional appearance of text fragments from the Bonferroni-corrected set of content features from the first three months (left), and the subsequent three months (right). Features with non-significant differences in the testing period are illustrated in grey

The performances of the classifiers that were constructed from connection-based source features were slightly better than classifiers from connection-based follower features. The accuracies of classifiers that selected from sources (86.2% to 88.0%) were slightly higher than their direct counterparts that were selected from followers (84.0% to 87.1%). The complete results are given in Table 2.

Table 2– The performances of classifiers trained to classify anti-vaccine tweets within the training period (the first three months)

Input Feature Set	Features selected	Precision	Recall	F1-score	Accuracy (95% CI)
Content: bigrams					
Bonferroni correction	11	0.74	0.56	0.63	82.0 (79.5-84.5)
p-value < 0.05	26	0.77	0.70	0.73	85.8 (83.5-88.1)
threshold = 3	37	0.88	0.74	0.82	89.8 (87.9-91.8)
Connections: followers					
Bonferroni correction	13	0.87	0.44	0.57	84.0 (81.6-86.4)
p-value < 0.05	21	0.89	0.48	0.62	85.5 (83.2-87.8)
threshold = 3	36	0.91	0.55	0.68	87.1 (84.9-89.3)
Connections: sources					
Bonferroni correction	18	0.88	0.55	0.67	86.7 (84.2-89.3)
p-value < 0.05	13	0.88	0.53	0.65	86.2 (83.9-88.5)
threshold = 3	28	0.88	0.60	0.71	88.0 (86.0-90.0)
Connections: followers, sources					
Bonferroni correction	23	0.86	0.57	0.68	87.0 (84.8-89.2)
p-value < 0.05	33	0.88	0.65	0.74	89.0 (86.9-91.1)
threshold = 3	39	0.88	0.67	0.76	89.4 (87.4-91.4)
Combined: bigrams, sources					
Bonferroni correction	17	0.86	0.63	0.72	87.8 (85.5-90.1)
p-value < 0.05	38	0.90	0.82	0.86	93.1 (91.3-94.9)
threshold = 3	42	0.91	0.84	0.87	93.8 (92.1-95.5)
Combined: bigrams, followers, sources					
Bonferroni correction	24	0.91	0.64	0.74	88.9 (86.7-91.1)
p-value < 0.05	51	0.94	0.84	0.88	94.4 (92.8-96.0)
threshold = 3	47	0.94	0.85	0.89	94.7 (93.1-96.3)

Classifier testing in period two

The performance of the classifiers was not sustained in the testing period, and the performance degradation observed from the training period to the testing period varied substantially across the classifiers (Table 3). The classifiers that included content features and had the highest accuracies in the training period exhibited the greatest degradation in performance when tested on tweets from the testing period.

Classifiers that used social connection information tended to perform similarly in the training and testing periods, with smaller changes in accuracy compared to the content-based classifiers (Table 3). These results are consistent with the statistical analysis of the features, which showed that the social connections were more consistently distributed across the two classes in the training and testing periods, compared to the text fragments.

Table 3 – The change in performance when applying the classifiers to the testing period (the subsequent three months)

Classifier	Accuracy (95% CI)	Accuracy change (%)
Bigrams (content)		
Bonferroni correction	85.2 (82.9-87.5)	3.2
p-value < 0.05	82.6 (80.1-85.1)	-3.2
threshold = 3	53.6 (50.4-56.9)	-36.2
Followers (connections)		
Bonferroni correction	86.0 (83.6-88.4)	2.0
p-value < 0.05	85.5 (83.1-87.9)	0.0
threshold = 3	84.1 (81.6-86.6)	-3.0
Sources (connections)		
Bonferroni correction	88.6 (86.4-90.8)	1.9
p-value < 0.05	81.6 (78.9-84.3)	-4.6
threshold = 3	87.3 (85.0-89.6)	-0.7
Bigrams, sources (both)		
Bonferroni correction	88.6 (86.4-90.8)	0.8
p-value < 0.05	82.2 (79.5-84.9)	-10.9
threshold = 3	87.1 (84.8-89.4)	-6.7

The two best performing classifiers were capable of distinguishing anti-vaccine tweets from all other tweets with 88.6% accuracy in the testing period. One was trained using only social connections and the other was trained using social connections and text fragments.

Discussion

We demonstrated that social connection information can be used to improve classifiers capable of identifying anti-vaccine opinions for HPV vaccines on Twitter, addressing the temporal degradation associated with using content features alone. While we have examined this phenomenon for only one topic, the results suggest that this approach may help to reduce the frequency at which social media surveillance systems would need to be updated through manual intervention.

Previous attempts at using social network information as features in supervised machine learning for Twitter classification have demonstrated reasonable performance – the best reported accuracies on various datasets were 68% and 73% using information from replies and retweets [30, 31], and between 58% and 77% using follower connections [32-34]. We believe our study is the first to demonstrate the difference in temporal degradation across classifiers constructed from content and social connection features.

The results suggest that information about who users follow, rather than who follows them, may be more useful for predicting the direction of their expressed opinions. A plausible explanation for this comes from the friendship paradox [35]. For any given user posting a tweet about HPV vaccines, the users they follow are likely to have more followers on average. More popular and influential users are expected to be better connected to the communities that tweet about HPV vaccines and as a consequence, may produce more useful features. The results may also suggest that there is a core of users that may be influential in vaccine information communities and that their followers tend to express similar opinions as a consequence of homophily or contagion [12, 14].

Limitations

There were several important limitations to the experiments reported here. Firstly, rate-limited access to Twitter precluded the instantaneous collection of user information each time we captured a relevant tweet, so calls to the API were staggered throughout the period and the information was collected only once for each user. However, given the stability of the social connections and the relatively small proportion of users that tweeted in both periods, this limitation is unlikely to have affected the conclusions. In addition, we did not apply any query expansion methods or evaluate the overall quality of the search terms we used.

Secondly, alternative feature space constructions and selection methods could have been chosen to produce classifiers, and these may have yielded different results. Specifically, there may be combinations of time-invariant text fragments that could out-perform our most accurate classifier (88.6% accuracy).

Finally – and importantly – we prospectively chose to demonstrate the results of the statistical analysis by implementing one type of classifier (SVMs using a radial basis function kernel) and fixed the parameter values to balance precision and recall in an unbalanced sample. If we had chosen other parameter values, different kernels, or other less appropriate machine learning algorithms, the results may have been different. However, since we tested the significant associations for all content and connection features independently of the classifier training and testing, our conclusions are largely independent of, but confirmed by, the construction of the classifiers.

Conclusion

For the task of classifying tweets about HPV vaccines as anti-vaccine or otherwise, information about the social connections between users provided a useful addition to the content of what people write. In particular, we showed that it is possible to use information about the users that people follow online to help predict their opinions. Our findings also suggest some potential avenues for the development of opinion forecasting – prospectively predicting the opinions of individuals and populations based on their social connections, rather than reactively classifying their opinions based on what they write.

References

[1] Burger JD, Henderson J, Kim G, and Zarrella G. Discriminating gender on Twitter, In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Edinburgh, United Kingdom, 2011; pp. 1301-9.

[2] Salathé M, and Khandelwal S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. PLoS Comput Biol 2011: 7: e1002199.

[3] Signorini A, Segre AM, and Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. During the influenza A H1N1 pandemic. PLoS ONE 2011: 6: e19467.
[4] Collier N, Son N, and Nguyen N. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011: 2(S9).
[5] Chew C, and Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE 2010: 5: e14118.
[6] Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, and Vespignani A. The Twitter of Babel: mapping world languages through microblogging platforms. PLoS ONE 2013: 8: e61981.
[7] Dodds PS, Harris KD, Kloumann IM, Bliss CA, and Danforth CM. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. PLoS ONE 2011: 6: e26752.
[8] Saif H, He Y, and Alani H. Alleviating data sparsity for Twitter sentiment analysis. In: 2nd Workshop on Making Sense of Microposts, 21st International Conference on the World Wide Web. Lyon, France, 2012; pp. 2-9.
[9] Zhang K, Cheng Y, Xie Y, Honbo D, Agrawal A, Palsetia D, Lee K, Liao W-K, and Choudary A. SES: Sentiment elicitation system for social media data. In: 11th International Conference on Data Mining Workshops. 2011; pp. 129-136.
[10] Coiera E. Social networks, social media, and social diseases. BMJ 2013: 346: f3007.
[11] Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011: 334: 1269-72.
[12] McPherson M, Smith-Lovin L, and Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociology 2001: 27: 415-44.
[13] Dietz K, Epidemics and rumours: a survey. J R Stat Soc 1967: 130: 505-28.
[14] Iyengar R, Van den Bulte C, and Valente TW. Opinion leadership and social contagion in new product diffusion. Marketing Science 2011: 30: 195-212.
[15] Coleman J, Katz E, and Menzel H. The diffusion of an innovation among physicians. Sociometry 1957: 20: 253-70.
[16] Leask J. Target the fence-sitters. Nature 2011: 473: 443-5.
[17] Gangarosa EJ, Galazka AM, Wolfe CR, Phillips LM, Gangarosa RE, Miller E, and Chen RT. Impact of anti-vaccine movements on pertussis control: the untold story. Lancet 1998: 351: 356-61.
[18] Fisher H, Trotter CL, Audrey S, MacDonald-Wallis K, and Hickman M. Inequalities in the uptake of human papillomavirus vaccination: A systematic review and meta-analysis. Int J Epidemiol 2013: 42: 896-908.
[19] Larson HJ, Smith D, Paterson P, Cumming M, Eckersberger E, Freifeld CC, Ghinai I, Jarrett C, Paushter L, Brownstein JS, and Madoff LC. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 2013: 13: 606-13.
[20] Madden K, Nan X, Briones R, and Waks L. Sorting through search results: a content analysis of HPV vaccine information online. Vaccine 2012: 30: 3741-6.
[21] Mahoney ML, Tang T, Ji K, and Ulrich-Schad J. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. JMIR Public Health Surveill 2015: 1: e2.
[22] Sotiriadis A, Dagklis T, Siamanta V, Chatzigeorgiou K, and Agorastos T. Increasing fear of adverse effects drops intention to vaccinate after the introduction of prophylactic hpv vaccine. Arch Gynecol Obstet 2012;285:1719-24.
[23] Joachims T. Transductive inference for text classification using support vector machines. In: Proceedings of the 16th International Conference on Machine Learning. Bled, Slovenia, 1999; pp. 200-9.
[24] Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv 2002: 34: 1-47.
[25] Pang B, Lee L, and Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing. Philadelphia, United States, 2002; pp. 79-86.
[26] Mullen T, and Collier N. Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing. Barcelona, Spain, 2004; pp. 412-8.
[27] Keerthi SS, and Lin C-J. Asymptotic behaviors of support vector machines with gaussian kernel. Neural Comput 2003: 15: 1667-89.
[28] Kohavi R, and John GH. Wrappers for feature subset selection. Artificial Intelligence 1997: 97: 273-324.
[29] Guyon I, and Elisseeff A. An introduction to variable and feature selection. J Machine Learning Res 2003: 3: 1157-82.
[30] Jiang L, Yu M, Zhou M, Liu X, and Zhao T. Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, United States, 2011; pp. 151-60.
[31] Boutet A, Kim H, and Yoneki E. What's in your tweets? I know who you supported in the UK 2010 general election. In: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media. Dublin, Ireland, 2012; pp. 411-4.
[32] Speriosu M, Sudan N, Upadhyay S, and Baldridge J. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the 1st Workshop on Unsupervised Learning in NLP, 16th Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, 2011; pp. 53-63.
[33] Hu X, Tang L, Tang J, and Liu H. Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. Rome, Italy, 2013; pp. 537-46.
[34] Tan C, Lee L, Tang J, Jiang L, Zhou M, and Li P. User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, United States, 2011; pp. 1397-1405.
[35] Feld SL. Why your friends have more friends than you do. Am J Sociol 1991: 96: 1464-77.

Address for correspondence

Xujuan Zhou and Adam G. Dunn, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 2109, NSW, Australia. Emails: susan.zhou@mq.edu.au; adam.dunn@mq.edu.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-766

Syndromic Surveillance of Infectious Diseases meets Molecular Epidemiology in a Workflow and Phylogeographic Application

Daniel Janies[a], Zachary Witter[a], Christian Gibson[a], Thomas Kraft[a], Izzet F. Senturk[b], Ümit Çatalyürek[b]

[a] *Department of Bioinformatics and Genomics, College of Computing and Informatics, University of North Carolina, Charlotte USA*
[b] *Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, USA*

Abstract

Traditionally, epidemiologists have counted cases and groups of symptoms. Modeling on these data consists of predicting expansion or contraction in the number of cases over time in epidemic curves or compartment models. Geography is considered a variable when these data are presented in choropleth maps. These approaches have significant drawbacks if the cases counted are not accurately diagnosed. For example, most regional public health authorities count influenza like illnesses (ILI). Cases of these diseases are designated as ILI if the patient exhibits fever, respiratory symptoms, and perhaps gastrointestinal symptoms. Several molecular epidemiological studies have shown that there are many pathogens that cause these symptoms and the relative proportions of these pathogens change over time and space. One way to bridge the gap between syndromic and genetic surveillance of infectious diseases is to compare signals of symptoms to pathogens recorded in molecular databases. We present a web-based workflow application that uses chief complaints found in the public Twitter feed as a syndromic surveillance tool and connects outbreak signals in these data to pathogens historically known to circulate in the same area. For the pathogen(s) of interest, we provide Genbank links to metadata and sequences in a workflow for phylogeographic analysis and visualization. The visualizations provide information on the geographic traffic of the spread of the pathogens and places that are hubs for their transport.

Keywords:

Infectious disease; Epidemiology; Foodborne Diseases; Public health surveillance; Twitter Messaging; Geo-referenced data; Genomics.

Introduction

Influenza-like illness (ILI) refers to a group of conditions such as fever, fatigue, cough, loss of appetite, aches, and sometimes gastrointestinal symptoms. The ILI concept remains a cornerstone of syndromic disease surveillance. However the pathogens that cause ILIs vary in space, time, taxonomy, and frequency and severity of infection. This gap between what syndromic surveillance purports to measure and what is being missed in terms of biology is striking. For example, in a retrospective study of 16,000 samples from patients clinically diagnosed with pandemic H1N1 influenza during April-July 2009 in Scotland, only 9% of the samples were tested positive via real-time PCR for pandemic H1N1 virus [1].

Similar studies of the molecular etiology have been conducted for respiratory tract infection in the Netherlands [2] (Table 1), United Kingdom [3] (Table 1), France [4], and China [5,6]. In the 2014, the United States had one of the first patients presenting the symptoms of fever and vomiting at an ER after a recent travel from West Africa and was sent home with antibiotics instead of being diagnosed for Ebola [7].

Table 1– A comparison of two similar studies illustrating temporal and geographic diversity of pathogens underlying lower respiratory tract diagnoses

NCBI taxid	Pathogen	Graffelman Netherlands	Creer UK
0	No pathogens detected	46.00	31.00
480	*Moraxella catarrhalis*	0.69	0.87
727	*Haemophilus influenzae*	6.92	4.37
810	*Chlamydia spp.*	1.38	0.00
1313	*Streptococcus pneumoniae*	4.85	13.10
2104	*Mycoplasma pneumoniae*	6.23	0.87
10535	Adenovirus	1.38	0.00
11118	Coronaviruses	0.00	4.37
11216	Parainfluenza virus type 3	2.08	0.00
11309	Influenza A and B pooled	0.00	16.59
197911	Influenza A	0.69	0.00
197912	Influenza B	24.92	0.00
12059	Rhinoviruses	1.38	22.71
12059	Enteroviruses	0.00	1.75
12730	Parainfluenza	0.00	2.62
12814	Respiratory syncytial virus	2.77	1.75
77133	Other bacteria	0.69	0.00
	Sum	100.00	100.00

Table 1 shows the percent of total pathogens underlying lower respiratory tract diagnoses identified by PCR and microbiological techniques by Graffelman and colleagues in the Netherlands (November 1998–June 2001) [2], and those identified by Creer and colleagues using real time PCR in the United Kingdom (May 2000-April 2001) [3]. Note that there are some differences in granularity of influenza testing and enterovirus testing.

We have created a system to: 1) bridge the gap between syndromic and genetic surveillance of infectious diseases and 2) provide information on the historical presence of pathogens associated with a cluster of related symptoms as identified in natural language from a public stream of a microblogging system, Twitter™.

Methods

We have created a live web-based workflow application (Figure 1) where analysts can obtain up-to-date syndromic data from a social media source. We correlate Tweet content to pathogens known to circulate in that geographic area using information such genetic data and metadata for pathogens of interest that has been published by subject matter experts.

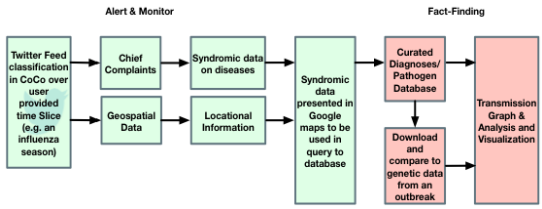


Figure 1. The data and analysis workflow consists of an alert and monitor phase focused on event modeling of a Twitter feed for clusters of Tweets with similar natural language corresponding to chief complaints.

Figure 1 shows that the rest of the workflow enables an analyst to perform fact finding by querying databases for sequence and metadata of pathogens that historically circulate in the locality of the Tweet cluster of interest. The analyst can download data from the links to Genbank (http://www.ncbi.nlm.nih.gov) to be combined with the fresh sequence and metadata from an out-break and/or can proceed with the transmission graph analysis and geographic visualization. The software modules implemented for the alert and monitor phase of the workflow is displayed in Figure 2.

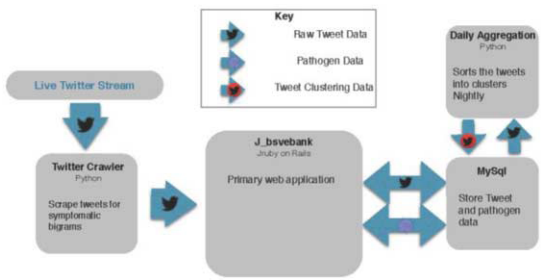


Figure 2- Software modules implemented for the alert and monitor part of the workflow. Arrows signify the flow of data from each module to the others in the system.

Tweets are gathered from the public stream, (https://dev.twitter.com/streaming/public) which serves a 1% sample of Tweets daily. Tweets are geocoded using user reported locations or after sending user profile data through Nominatim (http://nominatim.openstreetmap.org/) for matching geographical metadata to textual place names.

This work is done in python environment separate from the primary JRuby on Rails application to balance the load processing load. As Tweets are geocoded, they are sent to the Rails server. The Rails server receives the Tweets and stores them in a MySQL database where, once a night, they are indexed into Tweet clusters to be seen by the analyst the next morning or as a response to a retrospective query.

The natural language in Tweets is classified using the Bayesian chief complaint classifier (CoCo) of the Real-time Out-break and Disease Surveillance project (RODS) [8] to deter-mine potential relevance of Tweets by comparing their content to a list of 388 bigrams associated with infectious diseases (https://www.scribd.com/doc/250640730/terms-for-syndromic-surveillance).

Natural language terms from the Tweets of interest are then assigned to one of the seven classifications: hemorrhagic, neurological, rash, botulinic, respiratory, constitutional, gastrointestinal, or other, as determined by CoCo. These classifications allow us to identify and group Tweets in the light of their potential value as syndromic indicators. This information is stored in a MySQL database and can be searched by querying specific areas around the globe and slices of time.

In the next processing stage, CoCo and the geocoded Tweet clusters are associated with the pathogen datasets curated from NCBI's population genetic sequence database (aka "popset" http://www.ncbi.nlm.nih.gov/popset). These datasets contain orthologous genetic sequences, geographical information, temporal data, and host metadata from the strains of pathogens.

The data found in these related popsets are then subjected to multiple sequence alignment with MAFFT under the default settings [9]. The alignment is then subjected to phylogenetic tree search with TNT under the default settings [10]. At the visualization stage, a randomly chosen single phylogeny of the optimal set found is used to generate a transmission graph. The transmission graph is created by using the ancestor-descendent changes in the "place of isolation" character as optimized on the phylogeny. The transmission graph is assessed for betweenness and centrality, and displayed in Google Maps™ (GM) using in house code for phylogenetic visualization and mapping as described by Janies et al. [13].

Results

The application at work

At first view, an analyst is greeted with an interface dominated by a GM instance. An analyst can use search and filter tools located in separate sliding windows to select syndromic data that may be of interest or view all of the most recently recorded data. After selecting syndromic data the view focuses on the map where the relevant Tweet data is displayed and our "alert and monitor" process begins.

Our alert system is based on the concept of Tweet clusters. Tweet clusters share geography, collection date range, and natural language associated with the symptoms of infectious diseases. The number of Tweets in a cluster is recorded on a daily basis. This allows us to use historic Tweet cluster data in building a distribution for statistical analysis of incoming Tweets. When analyzing a new Tweet cluster, a collection of data is generated from the past clusters that were gathered over a period of three months.

For each bigram discovered within a Tweet we calculate the average number of times that bigram has appeared in the past three months, a standard deviation factor for the bi-gram's expected appearance, and a value representing how many times the bigram's appearance exceeds the standard deviation.

We use circles to visualize and represent Tweet clus-ters. The color of a circle indicates the intensity of the syndromic activity based on the statistical analysis and the circle size is proportional to the number of Tweets in the cluster. The circles are placed onto the GM at the respective geographic centers of the Tweet clusters (Figure 3).

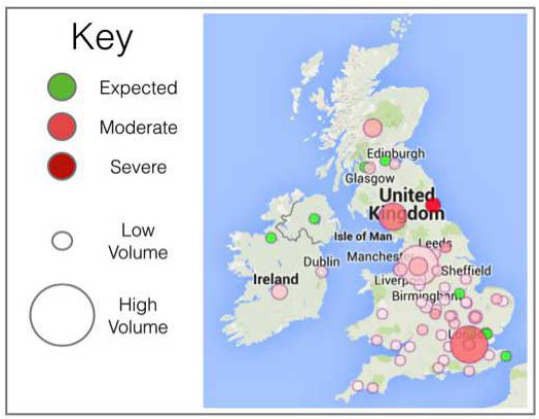


Figure 3- Tweet clusters from a winter day in 2015 the UK and Ireland.

Further analysis can be performed on a Tweet cluster by clicking on its circular icon on the GM. A window will drop down to display a series of statistics related to the chosen Tweet cluster in two separate tables. The first table displays the statistics describing cluster coloration within the alert system, while the other organizes the bigrams found in Tweets of the cluster into their respective CoCo classifications.

The first table informs the analyst whether it is worthwhile to to further investigate a Tweet cluster. This table displays the bigram of interest, its CoCo classification, the average number of times this bigram has appeared in the past 90 days, the standard deviation from the average, and the current appearance count. These values determine the color of the cluster in the alert system, and may influence users' interest to further explore the cluster.

Tweets are clustered by various CoCo classifications. The second table presents options to display more information about the classifications within a cluster. The count of Tweets associated with the CoCo classification is shown. To dig deeper, an analyst may view the actual Tweets or a pie chart summarizing the distribution of chief complaints (Figure 4).

For a cluster of interest, an analyst can query the population set database using CoCo diagnosis and locality information. Querying the pathogen population dataset returns a taxonomy pie chart of the pathogens in the area (Figure 5) and a series of scrollable datasets of genetic, geographic, and temporal metadata.

Each accession number within these datasets is linked to the National Center for Biotechnology Information's Genbank (http://www.ncbi.nlm.nih.gov). Using these links the analyst can leverage the information and tools in Genbank such as cross references to nucleotides, proteins, publications on the clinical or zoonotic aspects of the pathogen, basic local alignment to similar sequences, structural information, etcetera) (Figure 6).

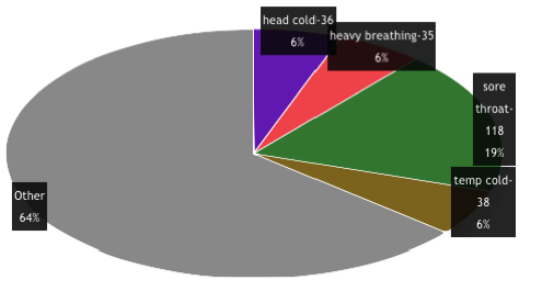


Figure 4- A view of Tweets associated with respiratory complaints in the United Kingdom in late 2014-early 2015.

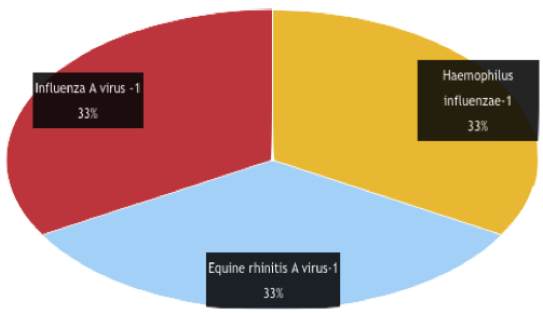


Figure 5- Pathogens associated with Tweets with respiratory complaints in the United Kingdom.

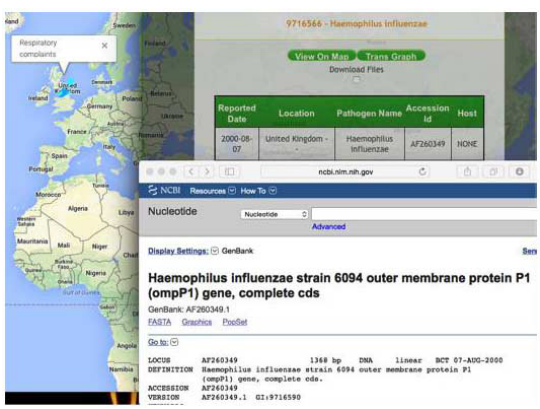


Figure 6. The analyst selected to view a Haemophilus influenzae genetic dataset associated with the classified Tweet in the United Kingdom.

For a pathogen population set of interest, the analyst can select to view its geographical diversity on GM and follow through with a phylogeographic visualization workflow. Our workflow does all the data wrangling including multiple sequence alignment, phylogenetic tree search, metadata optimization and visualization on the tree. The analyst can choose to view the results as an abstract and static transmission graph (Figure 7) or a literal and animated transmission map on GM (Figure 8). In Figure 7, A node represents geographic location for each isolate. The transmission graph reflects historical transmission links as viewed by the character evolution on a phylogeny. The size and darkness of the circles represent relative importance of a geographic place in spreading a pathogen in terms of a betweenness metric (e.g., the UK is important in this case)

[13]. The thickness of the lines and the size of ar-rows represent frequency of historical transmissions. In Figure 8, A transmission event is represented by a weighted link between two locations where the thickness of the link indicates the transmission frequency. The direction of the transmission is denoted by an animated arrow moving from transmission source to destination with a speed based on the transmission frequency (http://secure-basin-8938.herokuapp.com/medinfo.gif).

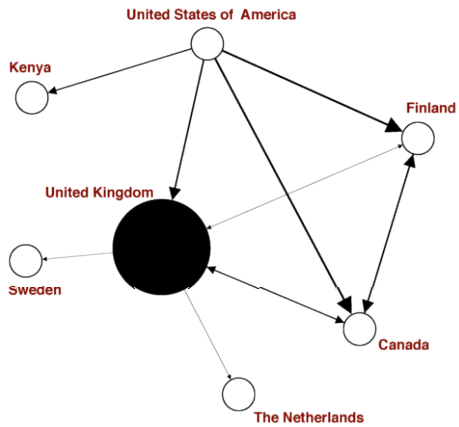


Figure 7. A transmission graph of Haemophilus influenzae isolates.

Figure 8. A screen shot from an animated transmission map of Haemophilus influenzae.

Discussion

We have created a bridge between syndromic and genetic surveillance of infectious diseases. Our system collects and classifies Tweets every day. Thus the results are dynamic. Between June and December of 2014 we classified over 1,000,000 Tweets and associated the Tweet clusters with corresponding pathogens selected from over 14,000 pathogen sequence datasets and metadata in the system. Siunce a human in the loop is important, the system is currently being prepared for usability testing with biosurveillance analysts.

The use case illustrated in the figures is derived from a study of molecular variations in light of vaccine design for non-typeable Hemophilus influenza [11]. The bacterium is a member of the Pasteurellaceae family and was once thought to be the cause of influenza. Hemophilus influenzae causes respiratory and other infections in humans. The worldwide spread of this pathogen is understudied, but the recent reports have shown that closely related strains are geographically diverse [12], which is consistent with our analysis.

The betweenness analysis can be carried out in conjunction with the social media analysis or on its own. The phylogenetic visualization modules of the system have also been used in the studies of H7 influenza [13] and Salmonella [14].

Conclusion

Syndromic surveillance is popular, as the data is growing rapidly and often widely shared. Given the current state of affairs, areas for improvement include the collection of more syndromic and genetic data. More data will provide denser information sets with which to connect syndromes and pathogens. Syndromic data can be expanded with the collection from sources other than Twitter and the including there in of multiple human languages. Next generation sequencing application in the public health settings is being widely adopted. Genetic data is often not rapidly shared but this practice is changing in some fields [15].

A major breakthrough will come when public health can rely primarily on genetic surveillance. Genetic surveillance can be used to accurately diagnose the pathogen underlying an outbreak. Accurate diagnosis is vital for treating patients well. Additionally, public health depends on appropriate use of antivirals and antibiotics to abate the evolution of the drug resistant organisms. As small portable low cost sequencing devices become available, we envision a future where genetic surveillance will be widespread and rapid. If these data are shared in a concerted fashion and there is support for development of tools to make sense of the data, then public health surveillance systems based on the biology of diseases will be within reach [16, 17]. If bio-surveillance analysts can devise tools to keep pace with the volume of these data and find new ways to work with it, we will be better able to anticipate the spread and flow of infectious diseases.

References

[1] Gunson R and Carman W. During the summer 2009 outbreak of "swine flu" in Scotland what respiratory pathogens were diagnosed as H1N1/2009? BMC Infectious Diseases 2011:11:192.

[2] Graffelman AW, Neven AK, Cessie S, Kroes ACM, Spring-er MP, and J van den Broek P. British Journal of General Prac-tice 2004:54:15–19.

[3] Creer DD, Dilworth JP, Gillespie SH, Johnston AR, Johnston SL, Ling C, Patel S, Sanderson G, Wallace PG, and McHugh TD. Aetiological role of viral and bacterial infections in acute adult lower respiratory tract infection (LRTI) in primary care. Thorax 2006:61:75–79.

[4] Schnepf N, Resche-Rigon M, Chaillon A, Scemla A, Gras G, Semoun O, Taboulet P, Molina JM, Simon F, Goudeau A, and LeGoff J. High Burden of Non-Influenza Viruses in Inflenza-Like Illness in the Early Weeks of H1N1v Epidemic in France. PLoS ONE 2011: 6(8): e23514.

[5] Yang ZF, Zhan YQ, Wang YT, Luo Y, Zhou H, Li JQ, Qin S, Guan WD, Liang ZW, Zhao SS, Huang QD, Mo ZY, Chen JX, and Chen RC. A pathogenic and clinical study of 882 cases of adult influenza-like illness in Guangzhou. 2010. Article in Chinese accessed abstract translated to English at PMID: 21176504.

[6] Li H, Wei Q, Tan A, and Wang L. Epidemiological analysis of respiratory viral etiology for influenza-like illness during 2010 in Zhuhai, China. Virology Journal 2013:10:143.

[7] Fernandez M and Sack K. Ebola Patient Sent Home Despite Fever, Records Show. New York Times, Oct. 10, 2014.

[8] Tsui FC, Espino J, Dato V, Gesterland P, Hutman J, and Wagner M. Technical Description of RODS: A Real-time Pub-lic Health Surveillance System. JAMIA 2003:10:5:399-408.

[9] Katoh S. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 2013:30:772-780.

[10] Goloboff P, Farris JS, and Nixon K. TNT, a free program for phylogenetic analysis. Cladistics 2008:24:5:774–786.

[11] Bolduc G, Bouchet V, Jiang R, Geisselsoder J, Truong-Bolduc C, Rice A, Pelton S. and Goldstein R. Variability of outer membrane protein P1 and its evaluation as a vaccine candidate against experimental otitis media due to nontypeable Haemophilus influenzae: an unambiguous, multifaceted approach. Infect. Immun. 2000: 68:8:4505-4517

[12] De Chiara M, Hood D, Muzzi A, Pickard D, Perkins T, Pizza M, Dougan G, Rappuoli R, Moxon E, Soriani R, and Donati C, Genome sequencing of disease and carriage isolates of nontypeable Haemophilus influenzae identifies discrete population structure. PNAS 2014: 111:14: 5439–5444.

[13] Janies DJ, Pomeroy LW, Krueger C, Zhang Y, Senturk I, Kaya K, and Çatalyürek UV. Phylogenetic visualization of the spread of H7 influenza A viruses. Cladistics Article first published online: 21 JAN 2015. http://onlinelibrary.wiley.com/doi/10.1111/cla.12107/abstract

[14] Hoffman M, Luo Y, Monday SR, Ottensen A, Gonzales-Escalona N, Timme R, Muruvanda R, Wang C, Kastanis G, Keys C, Janies D, Senturk I, Catalyurek UV, Hammack TS, Chu A, Myers R, Haendiges J, Wolfgang WJ, Schoonmaker-Bopp D, Evans P, Meng J, Strain E, Allard MW, and Brown EW. Origins of the Salmonella Bareilly strain causing a Foodborne Outbreak linked to Tuna from an Indian Facility. JID accepted.

[15] Janies DA, Voronkin IO, Das M, Hardman J, Treseder TW, and Studer J. Genome informatics of influenza A: from data sharing to shared analytical capabilities. Anim Health Res Rev. 2010. 11:1:73-9

[16] Janies D, Embi PJ, and Payne PR. Health-care hit or miss?: Collect genetic data on pathogens. Nature 2011:470:7334:329.

[17] Mounce, R and Janies D. Synergistic sharing of data and tools to enable team science. in Zander J and Mosterman PJ, eds. Computation for Humanity - Information Technology to Advance Society. Boca Raton, Florida: CRC press. 2013; pp. 435-448.

Address for correspondence

Daniel Janies, Ph.D.
Carol Grotnes Belk Distinguished Professor of Bioinformatics and Genomics
The College of Computing and Informatics
The University of North Carolina at Charlotte
9201 University City Boulevard
Charlotte, NC 28223 USA
djanies@uncc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-771

Fingerprinting Biomedical Terminologies – Automatic Classification and Visualization of Biomedical Vocabularies through UMLS Semantic Group Profiles

Bastien Rance[a], Thai Le[b], Olivier Bodenreider[c]

[a] *AP-HP, University Hospital Georges Pompidou; INSERM, UMR_S 1138, Centre de Recherche des Cordeliers, Paris, France*
[b] *Biomedical and Health Informatics, University of Washington, Seattle, WA, USA*
[c] *National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

Abstract

***Objectives**: To explore automatic methods for the classification of biomedical vocabularies based on their content. **Methods**: We create semantic group profiles for each source vocabulary in the UMLS and compare the vectors using a Euclidian distance. We explore several techniques for visualizing individual semantic group profiles and the entire distance matrix, including donut pie charts, heatmaps, dendrograms and networks. **Results**: We provide donut pie charts for individual source vocavularies, as well as a heatmap, dendrogram and network for a subset of 78 vocabularies from the UMLS. **Conclusions**: Our approach to fingerprinting biomedical terminologies is completely automated and can easily be applied to all source vocabularies in the UMLS, including upcoming versions of the UMLS. It supports the exploration, selection and comparison of the biomedical terminologies integrated into the UMLS. The visualizations are available at (http://mor.-nlm.nih.gov/pubs/supp/2015-medinfo-br/index.html).*

Keywords:

UMLS; Semantic Groups; Terminology Fingerprinting; Content-based Classification.

Introduction

The Unified Medical Language System® (UMLS) is a terminology integration system [1]. It provides broad coverage of the biomedical domain, from disorders to procedures to drugs to anatomical structures. While some source vocabularies focus on a subdomain of biomedicine (e.g., RxNorm for drugs), others, such as SNOMED CT and the NCI Thesaurus, provide coverage across biomedicine. However, selecting a biomedical terminology remains challenging for users, because there is no description of content coverage, i.e., no description of which subdomains are covered by a given terminology.

The UMLS used to provide a classification of source vocabularies based on usage. This classification, performed manually, leveraged the Medical Subject Headings (MeSH). This classification was heterogeneous, as it mixed usage and content categories. For example, categories such as "Nursing" and "Complementary Therapies" reflect usage, whereas the categories "Disease" and "Procedures" are based on content. Moreover, classification by usage does not necessarily align with classification by content. For example, the International Classification for Nursing Practice (ICNP®) and Nursing Interventions Classification (NIC) are both "Nursing" terminologies, although NIC predominantly contains therapeutic procedures, while ICNP also contains content about diagnoses and outcomes. In addition, source vocabularies may need to be classified into more than one category. Another limitation of this classification is that only the most frequently updated sources in the Metathesaurus were considered, because manual classification is labor-intensive. Overall, while useful to new users, this classification was imperfect and difficult to maintain for new versions of the UMLS in which new terminologies may have been introduced or modified significantly.

The objective of this work is to explore automatic methods for the classification of biomedical vocabularies based on their content. More specifically, we create a "fingerprint" (i.e., semantic profile) for each terminology in the UMLS by leveraging the categorization of UMLS concepts into semantic groups. These semantic group profiles form the basis for classifying and comparing the biomedical vocabularies based on their content. This classification could help terminology users to better understand the clinical or research purposes for which the terminologies are needed. Our approach is fully automatic, does not require any additional knowledge about the vocabularies, and can be easily deployed. We also explore several visualization techniques to render this classification. The semantic fingerprints we provide for biomedical terminologies could complement, if not replace, the classification of UMLS source vocabularies provided earlier.

Background

The Unified Medical Language System. The Unified Medical Language System® (UMLS) is assembled by integrating 179 source vocabularies. The UMLS Metathesaurus (version 2014AB) currently contains about 3.1 million concepts, i.e., clusters of synonymous terms coming from various source vocabularies. Each Metathesaurus concept is assigned at least one semantic type from the UMLS Semantic Network, a small network of 133 semantic types organized into a tree structure. The semantic types are partitioned into fifteen semantic groups (McCray et al. 2001), which represent broad subdomains of biomedicine, such as *Anatomy*, *Chemicals & Drugs*, and *Disorders*. Every semantic type is categorized into only one semantic group. The fifteen semantic groups are listed in Table 1, along with the number of Metathesaurus concepts in each group. In practice, the semantic groups provide a coarse categorization of the Metathesaurus concepts based on the principles of semantic validity, parsimony, completeness, exclusivity, naturalness, and utility. The semantic groups have been used in several applications, including visualizations of highly conceptual spaces [2], discovery of inconsistencies in the categorization of UMLS concepts [3], word-sense disambiguation [4], and quality assurance of value sets [5].

Visualization and cognition. We present different graphical representations of vocabularies based on their semantic content. There exists a broad body of literature describing the impact of visual displays on not only the speed of decision-making, but also its accuracy [6-9]. Cognitive theories provide context on the processes involved in visualizing information. This can range from the theories of Cleveland and McGill, who propose a set of elementary visual tasks for interpreting displays, to Pinker's models of cognitive processing from raw visual information to encoded visual descriptions [10, 11]. We also leverage some of the visualization techniques used for genomic datasets [12]. In the context of fingerprinting biomedical terminologies, visualization is an important component of presenting the information in a succinct manner to facilitate use by a broad range of stakeholders within the biomedical community.

Table 1– Distribution of Metathesaurus concepts by semantic groups

Semantic group name	Abbreviation	# concepts
Activities & Behaviors	ACTI	4,385
Anatomy	ANAT	122,298
Chemicals & Drugs	CHEM	813,426
Concepts & Ideas	CONC	48,711
Devices	DEVI	45,883
Disorders	DISO	544,829
Genes & Molecular Sequences	GENE	67,760
Geographic Areas	GEOG	4,426
Living Beings	LIVB	948,012
Objects	OBJC	16,175
Occupations	OCCU	1,506
Organizations	ORGA	2,220
Phenomena	PHEN	12,778
Physiology	PHYS	140,146
Procedures	PROC	374,195

Methods

Our method for classifying biomedical vocabularies based on their content can be summarized as follows. For each source vocabulary in the UMLS, we first create vectors reflecting the distribution of concepts among semantic groups (i.e., semantic group profiles). We then compare the semantic group profiles using a Euclidian distance. Finally, we apply several visualization techniques to the semantic group profiles.

Creating semantic group profiles

For each UMLS source vocabulary, we compute the frequency distribution of its concepts among the 15 semantic groups, which we record in a 15-dimensional vector. This is what we call the semantic group profile (or semantic fingerprint) of a source vocabulary. For example, SNOMED CT spans a variety of semantic groups, including *Disorders* (31%), *Chemicals & Drugs* (23%), *Procedures* (11%), *Anatomy* (7%) and *Devices* (3%). In contrast, 99% of the concepts from the Foundational Model of Anatomy (FMA) belong to the semantic group *Anatomy*. Its semantic profile is sparse, with few semantic groups other than *Anatomy* having a value other than 0. The set of vectors computed for each terminology forms a matrix of terminologies by semantic groups.

Comparing semantic group profiles

In order to compare two semantic group profiles, we use a Euclidian distance metric, that is, the straight line distance between two vectors, i.e., between two semantic group profiles. (We also tested other similarity metrics including cosine, Jaccard, and Dice. However, the Euclidian distance provided a range of values more suitable for defining groups of source vocabularies using hierarchical clustering.) We generate a distance matrix by calculating the Euclidian distance between terminologies pairwise.

We then use an agglomerative method of hierarchical clustering to group together similar semantic group profiles. The agglomerative hierarchical clustering algorithm starts with a distance matrix and identifies the pair of source vocabularies that are the most similar. This forms the first cluster. The distance matrix is then recalculated, with complete linkage defining the distance between clusters as the largest distance between any two of its elements. The elements of the matrix are compared to find the next closest pair between sources or clusters. This is repeated until a single agglomerative cluster of all source vocabularies is formed.

Visualizing semantic group profiles

We propose three different visualizations for the semantic group profiles depending on what we want to emphasize. Namely, we visualize single semantic fingerprints (i.e., single terminologies) with "donut" pie charts, sets of semantic fingerprints (i.e., multiple terminologies) with heatmaps, and associations between terminologies and semantic groups with network representations.

Visualizing single semantic group profiles – "donut" pie charts

We use ***"donut" pie charts*** for visualizing single semantic group profiles. In this visualization, the source is represented as a ring. The source ring contains arcs corresponding to each semantic group. The size of the arc is proportional to the proportion of the corresponding semantic groups in the source. In addition to displaying the profile of a give terminology, this representation also makes it easy to compare different profiles.

Visualizing sets of semantic group profiles – heatmaps and dendrograms

We provide a ***heatmap*** representation of the data found in the distance matrix. The source vocabularies are listed on the x-axis and the semantic groups are listed on the y-axis. Density on the heatmap corresponds to the percentage of all concepts within a source vocabulary that is found in a given semantic group. Density is color coded, with red for high percentages of a semantic group in terminology and yellow for low percentages. As a result, scanning a vertical slice (column) of the heatmap provides a visual representation of the semantic group profile for a source vocabulary. Conversely, by scanning a horizontal slice (row) of the heatmap, a user can easily identify those source vocabularies with a large proportion of concepts for this semantic group.

Additionally, we generate a ***dendrogram*** to visualize the hierarchical clustering of the source vocabularies. Short branches on the tree represent terminologies with similar semantic group profiles, while long branches represent more dissimilar source vocabularies. The dendrogram can be cut to obtain a given number of clusters. Clustering also helps contrast groups of terminologies with similar semantic group profiles within groups and different profiles across groups. An arbitrary number of clusters can be produced, depending on the threshold of similarity among clusters used.

Visualizing associations among terminologies – networks

In order to visualize associations among terminologies through semantic groups, we apply a bipartite network visualization to the semantic group profiles for visual display of content across multiple source vocabularies.

- Nodes represent semantic groups on the one hand and source vocabularies on the other.

- An edge from source S to semantic group G is drawn if the source vocabulary contains at least some percentage of concepts in G.

The network representation makes it easy to identify which source vocabularies share a high concentration of a particular semantic group, as these vocabularies all have edges to this semantic group. Different networks can be obtained by selecting different thresholds for the minimum proportion of concepts from semantic groups.

Implementation

All statistical analyses and heatmap visualization were performed using the R statistical software. Network display and "donut" pie chart leverage the JavaScript library "D3 for Data-Driven Documents".

Results

Visualizing single semantic group profiles – "donut" pie charts

Figure 1 shows the semantic group profiles of four UMLS source vocabularies. As mentioned earlier, the Foundational Model of Anatomy (FMA) contains almost exclusively anatomical concepts, displayed in light green. Similarly, the Online Mendelian Inheritance in Man (OMIM) vocabulary essentially contains gene (green) and disease (red) concepts. In contrast, SNOMED CT, a general clinical terminology, contains concepts from almost all semantic groups, with a large proportion of disease concepts. Finally, while the National Drug File-Reference Terminology (NDFRT) is a drug terminology, it also contains not only a majority of drug concepts (dark green), but also large numbers of concepts from other semantic groups, including *Disorders* (red) and *Physiology* (orange), because NDF-RT drugs are described drugs in terms of physiologic effect and mechanism of action (*Physiology*), as well as therapeutic intent (*Disorders*).

Visualizing sets of semantic group profiles – heatmaps and dendrograms

Figure 2 shows the heatmap and dendrogram resulting from the hierarchical clustering of 78 source vocabularies. (Although the distance matrix was computed for all source terminologies, the display is limited to these 78 vocabularies for readability. In practice, we filtered out non-English vocabularies. While translations of vocabularies contain new labels for concepts, their semantic content is identical to that of their English source. We also filtered vocabularies with fewer than 1,000 concepts since their small size limits their overall significance.)

Columns from the heatmap represent the semantic group profiles of individual source vocabularies. For example, the Foundational Model of Anatomy is represented by a single red spot for the semantic group *Anatomy*, while SNOMED CT spans multiple semantic groups in the column. Conversely, the rows of the heatmap reflect the density in concepts from a given semantic group. The large red bar in the lower right corner corresponds to a high density of concepts from the *Disorders* semantic group in disease terminologies.

The clustering algorithm was (arbitrarily) required to produce 6 clusters. Each cluster is rooted by the top subdivisions of the dendrogram (and highlighted by boxes with solid lines on the figure). Clusters range in size from 1 source vocabulary (HGNC) for the leftmost cluster, to 26 source vocabularies for the rightmost cluster. Some clusters are homogenous. For example, cluster 1 contains one gene terminology, cluster 2 contains six procedure terminologies and cluster 4 contains two terminologies primarily containing organisms. In contrast, the remaining clusters are heterogeneous and subgroups can easily be identified within them. For example, the large cluster 3 groups drug terminologies such as RxNorm, device terminologies, such as the Current Procedural Terminology (CPT), and general terminologies, such as SNOMED CT. Similarly, cluster 5 groups organism terminologies, such as the NCBI Taxonomy, anatomical terminologies, such as the Foundational Model of Anatomy (FMA), administrative terminologies, such as the HL7 value sets (HL7V2.5), and terminologies with focus on physiological concepts, such as LOINC and the International Classification of Functioning (ICF). Finally, cluster 6 clearly groups disease terminologies, some of which contain only disease concepts (e.g., ICD 10-CM) and other contain concepts from another group (e.g., genes and diseases in OMIM).

Visualizing associations among terminologies – networks

The bipartite network we created for visualizing associations among terminologies contains two types of nodes. The source vocabularies are represented in green, while the semantic groups are in yellow. Edges are drawn between a source vocabulary and a semantic group if the vocabulary contains at least 5% of concepts from this semantic group. (This arbitrary threshold can be modified to reflect stronger associations.) In Figure 3, source vocabularies that contain at east 5% of concepts from the Disorders semantic group are highlighted. Similarly, as shown in the inset from Figure 3, it is also possible to highlight all semantic groups for a given source vocabulary (i.e., all the semantic groups, whose concepts constitute at least 5% of the source vocabulary).

Discussion

Use cases and applications

The semantic group profiles provide a method for assessing the similarity among source vocabularies in the UMLS. This general technique can be applied to terminology exploration, terminology selection and terminology comparison.

Exploring terminologies

Novice users of the UMLS sometimes have difficulties grasping the differences among the many source vocabularies in the Metathesaurus. While the UMLS Terminology Services browser allows users to find the details about individual Metathesaurus concepts and their relations, it does not provide an overview of sets of concepts in source vocabularies. Our donut pie charts, heatmap and network visualizations provide an overview of the content of the source vocabularies. More specifically, they provide a coarse description of the semantics of these terminologies, making it possible to quickly identify the major semantic areas in a given vocabulary.

Selecting terminologies

One common use case is to select the best terminology for a given application. For example, if an application requires disease concepts, our visualizations make it easier for a user to identify candidate terminologies, i.e., terminologies containing a large proportion of concepts from the semantic group *Disorders*. In practice, a user will look for a large red arc on the donut pie charts, or might scan the *DISO* row on the heatmap, looking for red spots. Alternatively, our user could also select the *DISO* node on the network visualization and explore all source vocabularies linked to it, having set an appropriate threshold for the minimal proportion of concepts from this semantic group required for edges to be drawn.

Comparing terminologies

The heatmap is also the visualization of choice for analyzing sets of source vocabularies, especially after the hierarchical

clustering has grouped together those terminologies that have similar semantic group profiles. The clusters displayed in Figure 2 and presented in the Results section are relatively easy to interpret, with minimal prior knowledge of the terminologies themselves. Similarity clusters can also be quantified, since the basis for clustering is the Euclidian distance computed among the semantic group profiles for individual source vocabularies.

Content-based vs. usage-based classification

Our work was motivated in part by the limitations of the usage-based classification the UMLS documentation used to provide. We compared the two classification approaches for the 55 source vocabularies for which it was available. The general trend is that there is limited overlap between the two classifications. Categories from the usage-based classification are generally associated with several semantic groups, and a given semantic group is generally associated with multiple categories from the usage-based classification, with no obvious patterns in these associations. One exception is the usage-based category "Adverse drug reaction reporting" that contained only one source vocabulary (MedDRA) and majoritarily contains concepts from the semantic group disorders. In fact, the two classifications provide different views on the source vocabularies and are complemantary. For example, it would be impossible to identify consumer health vocabularies or nursing vocabularies simply from the semantic group profiles. However, as mentioned earlier, unlike the manual usage-based classification, our semantic group profiles can be applied automatically to any new version of the UMLS. Finally, another advantage of our method is that, because it is a vector-based representation of the source vocabularies, it lends itself nicely to visual representation.

Limitations

Many concepts have more than one semantic type; however, these multiple semantic types are generally categorized into the same semantic group. Therefore most concepts are categorized by only one semantic group. In fact, only about 1,000 concepts have multiple semantic groups. As a result, the fifteen semantic groups form partition for over 99.9% of all UMLS concepts, and are thus virtually disjoint.

For the purpose of computing the distribution of the concepts from a source vocabulary into semantic groups, the concepts that have multiple semantic groups should logically not be counted more than once. In practice, these concepts are so few in the UMLS that the effect of double-counting them has no significant effect on the frequency distributions.

The assignment of a semantic type to a UMLS concept is sometimes subjective and can be arguable. Many concepts are categorized with multiple semantic types. In contrast, all UMLS concepts are categorized in 15 disjoint semantic groups. Because the semantic groups are broader, the assignment of concept to a group is less likely to be arguable. However, some groups can be viewed as too general for this application. For example, the semantic group "Chemicals" contains both drugs and other chemicals. A user could be interested in retrieving drug vocabularies, rather all chemical vocabularies. As suggested in [13], the grouping of semantic types into semantic groups could be modified to fit the requirements of a particular application.

Future work

In this study we have used our fingerprinting methodology on UMLS terminologies. Leveraging concept mappings among terminologies, our approach could be used to automatically classifiy the content of terminologies in other repositories such as the NCBO Bioportal. The semantic group-based fingerprint could also be used in a broader context, for example to help characterize the content of biomedical articles or clinical texts.

Conclusion

The growth of the UMLS makes it difficult for users to select appropriate source vocabularies for a given purpose. In this article, we present a new method to classify biomedical source vocabularies based on their content. We leverage the high level semantic categorization of concepts in semantic groups to create a profile for each source vocabulary. Our approach is completely automated and can easily be applied to all source vocabularies in the UMLS, including upcoming versions of the UMLS.

To assist the user in the exploration of available source vocabularies, we propose several visualizations reflecting the individual content of source vocabularies (donut pie charts, heatmaps), as well as the relations among source vocabularies (dendrogram, network). We are currently collaborating with the UMLS team to add the graphical representations to the UMLS documentation, as a complement to the classification they already provide.

References

[1] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;32(Database issue):D267-70.

[2] Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. *J Biomed Inform* 2003;36(6):414-32.

[3] Mougin F, Bodenreider O, Burgun A. Analyzing polysemous concepts from a clinical perspective: application to auditing concept categorization in the UMLS. *J Biomed Inform* 2009;42(3):440-51 http://www.ncbi.nlm.nih.gov/pubmed/19303057.

[4] Jimeno-Yepes A, McInnes BT, Aronson AR. Collocation analysis for UMLS knowledge-based word sense disambiguation. *BMC Bioinformatics* 2011;12 Suppl 3:S4 http://www.ncbi.nlm.nih.gov/pubmed/21658291.

[5] Jiang G, Solbrig HR, Chute CG. Quality evaluation of value sets from cancer study common data elements using the UMLS semantic groups. *J Am Med Inform Assoc* 2012;19(e1):e129-36 http://www.ncbi.nlm.nih.gov/pubmed/22511016.

[6] Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ* 1999;318(7197):1527-31.

[7] Feldman-Stewart D, Brundage MD, Zotov V. Further insight into the perception of quantitative information: judgments of gist in treatment decisions. *Med Decis Making* 2007;27(1):34-43.

[8] Hoeke JO, Bonke B, van Strik R, Gelsema ES. Evaluation of techniques for the presentation of laboratory data: support of pattern recognition. *Methods Inf Med* 2000;39(1):88-92.

[9] Morrow DG, Hier CM, Menard WE, Leirer VO. Icons improve older and younger adults' comprehension of medication information. *J Gerontol B Psychol Sci Soc Sci* 1998;53(4):P240-54.

[10] Cleveland W, McGill R. Graphical perception: Theory, experimentation, and application to the development of

graphical methods. *Journal of the American statistical association* 1984;79(387):531-554.

[11] Pinker S. A theory of graph comprehension. In: Friedle R, editor. *Artificial intelligence and the future of testing*. Hillsdale, NJ: Erlbaum; 1990.

[12] Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. *Genome Med* 2013;5(1):9 http://www.ncbi.nlm.nih.gov/pubmed/23363777.

[13] McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. *Stud Health Technol Inform* 2001;84(Pt 1):216-20.

Acknowledgments

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine. BR is supported in part by the the Cancer Research and Personalized Medicine - CARPEM project (Cancer Research Site - SIRIC). The authors thank Steve Emrick for useful discussions and feedback on the semantic fingerprints of UMLS source vocabularies.

Address for correspondence

Corresponding author: olivier@nlm.nih.gov

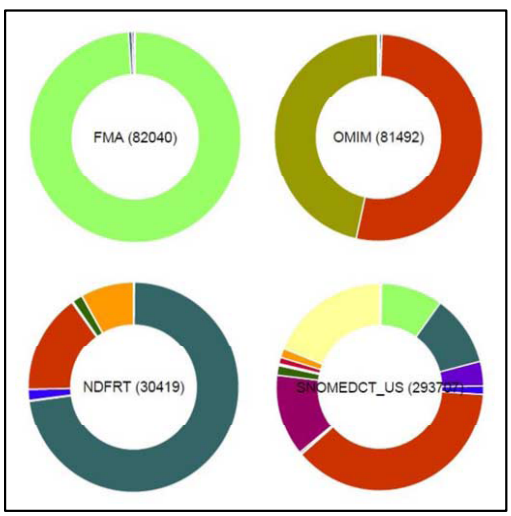


Figure 1– "Donut" pie charts for 4 UMLS source vocabularies. Color code: anatomical concepts (light green), gene (green) diseases (red), drugs (dark green), physiology (orange), leaving being (magenta), procedure (light yellow)

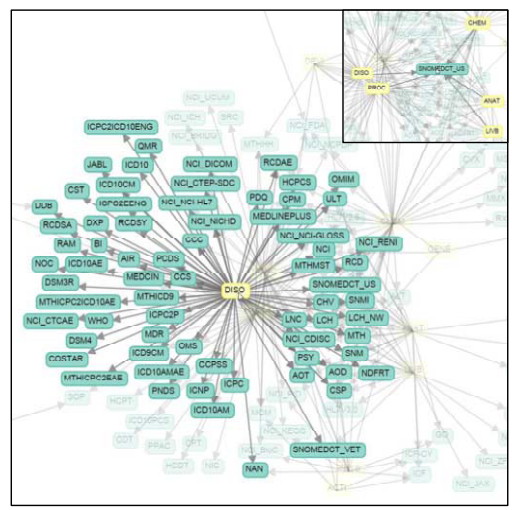


Figure 3– Network visualization of UMLS source vocabularies (green) linked to the semantic group Disorders (yellow) [Inset: Network visualization of SNOMED CT and its associations with several semantic groups]

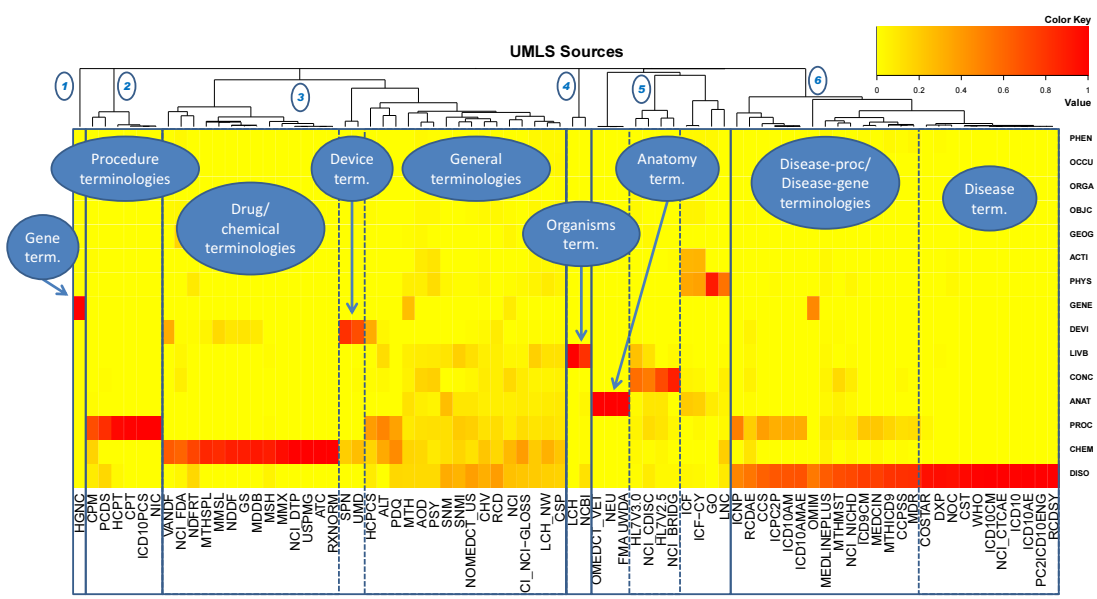


Figure 2– Heatmap of the UMLS terminologies and semantic groups. Bright yellow defines the absence of a semantic group in a terminology. On the opposite, bright red denotes a high percentage of concept of the corresponding semantic group in the terminology

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-776

Harmonizing Nursing Terminologies

Amy Coenen[a], Fernanda Paese[b], Virginia Saba[c], Kay Jansen[a], Nicholas R. Hardiker[d], Tae Youn Kim[e]

[a] *College of Nursing, University of Wisconsin-Milwaukee, Milwaukee, WI, USA*
[b] *Department of Nursing, Federal University of Santa Catarina, Florianópolis, SC, Brazil*
[c] *School of Nursing, Georgetown University, Washington DC, USA*
[d] *School of Nursing, Midwifery, Social Work & Social Sciences, University of Salford, Manchester, UK*
[e] *Betty Irene Moore School of Nursing, University of California Davis, Sacramento, CA, USA*

Abstract

In this paper, the authors report on a study aimed at harmonising two nursing terminologies, the Clinical Care Classification (CCC) and the International Classification for Nursing Practice (ICNP®). As the electronic health record evolves and the need for interoperability extends beyond local and national borders, a degree of standardisation across healthcare terminologies become essential. Harmonising across terminologies results in a) increased consensus relating to domain content and b) improvements in the terminologies involved. Findings from this study suggest that there is much overlap of content in nursing terminologies. The continued harmonisation between nursing terminologies and other healthcare terminologies are recommended to achieve international interoperability.

Keywords:

International Classification for Nursing Practice; Nursing; Health Terminology Standards; Interoperability.

Introduction

The International Council of Nurses (ICN) has supported the ongoing development and maintenance of the International Classification for Nursing Practice (ICNP) for more than 25 years. An early aim of the ICNP Programme was to harmonize nursing and healthcare terminologies to facilitate interoperability in the electronic health records (EHR). The purpose of this study was to initiate harmonisation of the Clinical Care Classification (CCC) System and ICNP with a focus on nursing problems or diagnoses. The source terminology comprised nursing problems in CCC (n=176). The target terminology comprised ICNP problems or negatively judged nursing diagnoses (n=491).

Background

ICN has supported the ICNP since 1989. ICNP is a standardized terminology, available in 17 spoken languages, comprising statements that represent nursing diagnoses, outcomes, and interventions. It is recognized within the WHO Family of International Classifications as a Related Classification and work is underway to harmonize ICNP with SNOMED CT through a Collaboration Agreement with International Health Terminology Standards Development Organisation (IHTSDO) [1]. At the heart of ICNP is an ontology, represented in the Web Ontology Language (OWL) [1]. This facilitates cross-mapping while allowing delivery in a range of formats to suit the needs of its users.

A second Collaboration Agreement [3] guides harmonisation work with the Clinical Care Classification (CCC), a terminology also focused on nursing and designed specifically for computer-based documentation and processing [4]. In common with ICNP, CCC is compositional in nature (particularly with respect to nursing interventions) and covers nursing diagnoses, outcomes and interventions. These similarities may facilitate comparison; they may also add complexity in harmonising the two terminologies.

In addition to its role as a formal reference terminology, ICNP also is often used as an interface terminology, in common with CCC. The diagram below (Figure 1) displays some similarities and differences in the ways CCC and ICNP concepts are organised and highlights the areas of attention in this study: nursing problems (negative diagnoses). In CCC all nursing diagnoses are judged negatively (as problems). In ICNP, nursing diagnoses may be judged as either negative (e.g. pain) or positive (no pain). ICNP positive nursing diagnoses can be used to identify assets, goals or outcomes. Because CCC nursing diagnoses are patient problems, the focus in this study was on ICNP negative nursing diagnoses.

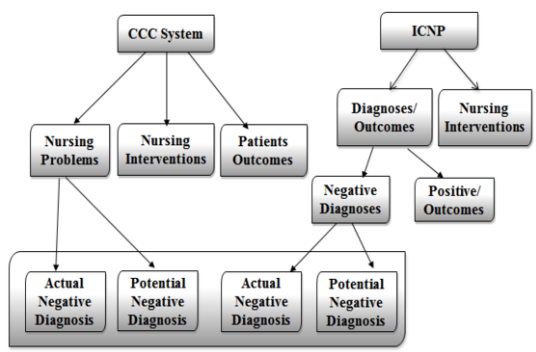


Figure 1 - Structural comparison of CCC and ICNP

Methods

In the mapping process, CCC nursing problems (Version 2.5) [4] acted as the source terms and ICNP nursing diagnoses (2013 release) [5] were the targets. A total of 176 CCC

nursing problems were examined for equivalent concepts in ICNP using a combination of automated and manual (human) processes. The Unified Medical Language System (UMLS) which is a terminology resource maintained by the United States (US) National Library of Medicine (NLM) was used to develop candidate equivalencies, using similar techniques to those described in [6]. Two experts in ICNP and CCC reviewed the candidates independently and compared results. The research team members revised the two sets of equivalencies guided by the CCC textual definitions and ICNP formal definitions. Revisions were made until 100% agreement was achieved among the experts.

Results

For each CCC problem concept an equivalent ICNP concept was identified. Complete coverage of CCC problems in ICNP indicates overlap between the two terminologies. However, not all mappings were one-to-one. Each CCC problem concepts (n=176) mapped to one of 169 equivalent ICNP concepts, which suggests a degree of redundancy in CCC. Specific findings in the equivalency mapping are reported here: (a) one-to-one mappings, (b) one-to-many mappings, (c) actual and potential (risk) problems, (d) use of positive ICNP diagnosis i.e. diagnoses that would be considered 'normal', (e) use of lexical and semantic mappings, and (f) similarities and differences between the two terminologies.

One-to-One Mappings

All but six CCC nursing problems (n=163) had a one-to-one mapping with an ICNP concept. Examples are provided in Table 1.

Table 1 - One-to-One mapping between CCC and ICNP®

CCC Diagnosis	CCC Code	ICNP Diagnosis	ICNP Code
Fall Risk	N33.6	Risk for Fall	10015122
Verbal Impairment	M28.1	Impaired Verbal Communication	10025104
Family Process Alteration	M29.0	Impaired Family Process	10023078
Tobacco Use	N58.1	Tobacco Use	10022247

Many-to-One Mappings

In six cases, more than one CCC concept mapped onto the same ICNP nursing diagnosis. For example, both CCC "Skin Integrity Alteration" (R46.0) and CCC "Skin Integrity Impairment (R46.2) mapped onto the ICNP concept "Impaired Skin Integrity" (10001290) (see Table 2). In these cases, the reviewers considered more than one CCC concept to be semantically equivalent to an ICNP concept, which suggests redundancy among CCC concepts or perhaps a need to revise CCC textual definitions for these six concepts.

Actual problems and potential (risk for) problems

Both CCC and ICNP included judgments about nursing diagnoses regarding potentiality of a problem. All concepts across both terminologies were judged as being either actual or potential (at risk).

Of the 176 CCC nursing diagnoses, 28 (5.7%) might be considered to be a potential problem (or risk) (see Figure 2). Examples of CCC risk diagnoses included: "Risk for Fall" and "Suicide Risk". ICNP includes these concepts among its much larger number of 113 potential problems (risks), for example "Risk for Complications during Pregnancy".

Table 2 - Many-to-One mappings between CCC and ICNP

CCC Nursing Diagnoses (13 of 176)	ICNP Nursing Diagnoses (6 of 169)
A01.0 – Activity Alteration A01.1 – Activity Intolerance	10000431 – Activity Intolerance
A01.5 – Physical Mobility Impairment Q44.3 – Kinesthetic Alteration	10001219 – Impaired Mobility
E11.0 – Family Coping Impairment E11.2 – Disabled Family Coping	10034789 – Impaired Family Coping
M32.0 – Socialization Alteration M32.1 – Social Interaction Alteration	10001022 – Impaired Socialization
038.0 – Self-Care Deficit 038.1 – Activities of Daily Living Alteration 038.2 – Instrumental Activities of Daily Living Alteration	10023410 – Self Care Deficit
R46.0 – Skin Integrity Alteration R46.2 – Skin Integrity Impairment	10001290 – Impaired Skin Integrity

Use of a non-problem (positive) ICNP diagnosis

Among the 169 ICNP nursing diagnoses mapped with CCC problems, a concept "Grief" (10022345) was considered a positive nursing diagnosis/outcome in ICNP but a problem (or negative) nursing diagnosis in CCC. That is, although the CCC concept "Grief" (E53.0) was lexically equivalent to the ICNP concept, the semantic meanings of the concepts were considered to be different based on 'judgment state" of each concept. The two terminology systems defined the 'judgment state' of the concept, "Grief", differently. In the CCC terminology all nursing diagnoses are considered to be problems. In ICNP nursing diagnoses and outcomes (often also used as patient goals) are modeled similarly and each includes a judgment; either positive or negative. In ICNP, "Grief" is judged as normal (or positive): a normal emotion.

Use of lexical and semantic mappings

In only one case it was determined that CCC had a concept with a similar lexical expression as ICNP but was actually mapped to a different concept. Both CCC and ICNP had the concepts "Kinesthetic Alteration" (Q44.3) and "Impaired Kinaesthesia" (10034886). The CCC concept "Kinesthetic Alteration" was mapped to the ICNP concept "Impaired Mobility" (rather than the ICNP concept "Impaired Kinaesthesia"). Again, this mapping was based on semantic definitions of concepts in each terminology.

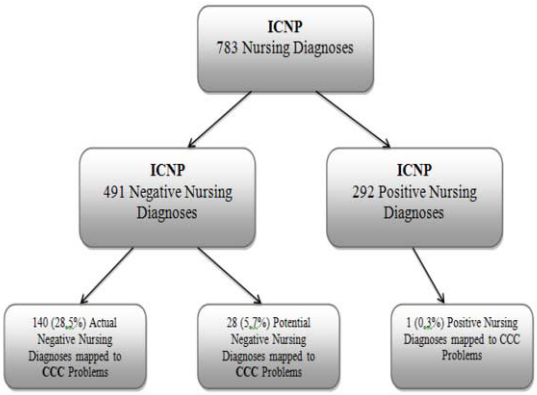


Figure 2 - Results of mappings of nursing diagnoses between CCC and ICNP

Within this study, 25% (44) of the source nursing concepts had identical lexical expressions in the target, 12% (21) had similar lexical expressions and 63% (111) had different lexical expressions, but exact semantic meaning (see Table 3).

Table 3 - Examples of lexical similarities and differences for semantically-equivalent concepts

	CCC Diagnosis	CCC Code	ICNP Diagnosis	ICNP Code
Lexically identical	Fatigue	A01.4	Fatigue	10000695
Lexically similar	Blood Pressure Alteration	C06.1	Altered Blood Pressure	10022954
Lexically different	Spiritual State Alteration	E14.0	Impaired Spiritual Status	10023336

Similarities and difference between two terminologies

Both CCC and ICNP provide a conceptual representation of nursing diagnoses. A major difference in the two terminologies is in the representation of definitions or descriptions of concepts. ICNP is founded on a formal OWL ontology. Definitions for ICNP concepts are represented formally in terms of hierarchical placement and relationships with other concepts. In contrast, each CCC concept has a textual description, but no formal definitions.

Another major difference is that ICNP includes a greater number of nursing diagnoses (n=783) than CCC (n=176). CCC nursing problems are divided between actual negative diagnoses and potential negative diagnoses. ICNP has these, but also includes positive (i.e., 'normal') diagnoses.

A further difference between the two terminologies is that CCC concepts are organised in a conceptual framework. The CCC framework includes 21 Care Components, for example Fluid Volume component (which includes the diagnosis "Electrolyte Imbalance") and Health Behavior Component (which includes the diagnoses of "Noncompliance of Medication Regimen" and "Home Maintenance Alteration"). These CCC Care Components are mutually exclusive, and each of the 176 CCC problems is placed in only one component. The ontological nature and modeling of ICNP allows for multiple parents within the ICNP hierarchy.

These differences present challenges for semantic mapping.

Discussion and Conclusion

This study examined the equivalency mapping between CCC and ICNP negative diagnoses (or problems) concepts. This project moves us closer to a broader harmonisation of nursing content. The mapping of concepts between CCC and ICNP is important in assuring interoperability of data and information across nursing care specialties and settings.

In addition to supporting interoperability among nursing terminologies, ICN also is formally collaborating with IHTSDO to harmonize ICNP and SNOMED CT. As the ICNP is mapped to SNOMED-CT, ICNP-SNOMED CT nursing equivalency tables can be used (together with ICNP-CCC equivalency tables) to automatically generate equivalencies between CCC and SNOMED CT, which will be reported elsewhere. The ongoing harmonisation project between ICNP and SNOMED CT will enhance further interoperability across nursing terminologies and reduce the burden of maintaining multiple, independent equivalency tables.

In addition to contributing to interoperability between nursing and other health terminologies, the findings of this study provided the opportunity to re-examine both CCC and ICNP and make improvements. Specific changes were made to ICNP based on the findings from this study. For example, changes were made to preferred terms in ICNP: "Impaired Ability to Feel" (10022619) was changed to "Impaired Tactile Perception"; "Impaired Ability to See" (10022748) was changed to "Impaired Vision"; and "Depression" (10022402) was changed to "Depressed Mood". These changes better represent both the formal definition and English lexical expression of concepts.

The finding about the concept "Grieving" or "Grief" raises important issues. A complex concept such as "Grieving", may have different judgment states (e.g. normal, risk for problem), depending on the way that the concept is perceived. Leveraging the use of standard terminologies and ontologies in clinical practice requires clarity about the meaning of such concepts.

Both of the terminologies in this study had nursing diagnostic concepts to represent actual problems and potential (risk for) problems. Risk factor identification is especially relevant for nursing practice targeted at health promotion and disease prevention. It is recommended that the significant number of potential problems or risks in ICNP continue to be reviewed and further concepts added based on use cases from nursing practice.

This project aimed to harmonize CCC problems and ICNP nursing diagnostic concepts. While the direction of mapping within the study was from CCC as the source to ICNP as the target, it is interesting to note that the majority (66.8%) of ICNP concepts have no equivalent in CCC. Due to the large number of unmapped ICNP negative nursing diagnoses (n = 323) a new study is underway to examine allocations of these ICNP concepts to the CCC 21 Care Components. Further review of the assignments between ICNP and the 21 Care Components will explore the theoretical organisation of the nursing domain content using the CCC framework [4].

The equivalency table resulting from this project facilitates the interoperability of CCC and ICNP for data exchange and comparison. In addition, CCC data can now be compared to other nursing and healthcare terminologies already harmonized with ICNP.

Acknowledgment

The authors acknowledge support from the International Council of Nurses.

References

[1] International Health Terminology Standards Development Organisation (IHTSDO). ICN and IHTSDO Team-up to Ensure a Common Health Terminology. (March, 2010). IHTSDO, 2010. Available from:

http://www.ihtsdo.org/resource/resource/72

[2] Hardiker NR, and Coenen, A. Interpretation of an international terminology standard in the development of a logic-based compositional terminology. Int J Med Inform 2007: 76(2): 274-80.

[3] International Council of Nurses (ICN). Harmonising nursing terminology internationally: Collaboration between the International Council of Nurses and SabaCare. (June, 2012). ICN, 2012. Available from: http://www.icn.ch/images/stories/documents/news/press_releases/2012_PR_09_ICNP_CCC.pdf

[4] Saba VK. Clinical Care Classification (CCC) System Version 2.5 User's Guide, 2nd ed. New York (NY): Springer Publishing Company, 2012.

[5] International Council of Nurses (ICN). International Classification for Nursing Practice Version 2. [Internet], ICN, 2013. Available from: http://www.icn.ch/images/stories/documents/pillars/Practice/icnp/translations/icnp-English%20translation.pdf.

[6] Kim, TY, Hardiker, NR, and Coenen, A. Inter-terminology mapping of nursing problems. Journal of Biomedical Informatics 2014: 49: 213–20.

Address for correspondence

Amy Coenen, College of Nursing, University of Wisconsin – Milwaukee, 1921 E. Hartford Avenue, Milwaukee, WI 53211, USA.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-780

Developing a National-Level Concept Dictionary for EHR Implementations in Kenya

Aggrey Keny[a], Steven Wanyee[b], Daniel Kwaro[c], Edwin Mulwa[d], Martin C. Were[a,e]

[a] *Academic Model Proving Access to Healthcare(AMPATH), Eldoret, Kenya*
[b] *International Training and Education Centre (ITECH)-Kenya, Nairobi, Kenya*
[c] *The Kenya Medical Research Institute / Centers for Disease Control (KEMRI-CDC), Nairobi, Kenya*
[d] *KEMRI-UCSF Family AIDS Care and Education Services (FACES), Kisumu, Kenya*
[e] *Indiana University School of Medicine and Regenstrief Institute, Inc., Indianapolis, IN, US*

Abstract

The increasing adoption of Electronic Health Records (EHR) by developing countries comes with the need to develop common terminology standards to assure semantic interoperability. In Kenya, where the Ministry of Health has rolled out an EHR at 646 sites, several challenges have emerged including variable dictionaries across implementations, inability to easily share data across systems, lack of expertise in dictionary management, lack of central coordination and custody of a terminology service, inadequately defined policies and processes, insufficient infrastructure, among others.

A Concept Working Group was constituted to address these challenges. The country settled on a common Kenya data dictionary, initially derived as a subset of the Columbia International eHealth Laboratory (CIEL) / Millennium Villages Project (MVP) dictionary. The initial dictionary scope largely focuses on clinical needs. Processes and policies around dictionary management are being guided by the framework developed by Bakhshi-Raiez et al. Technical and infrastructure-based approaches are also underway to streamline workflow for dictionary management and distribution across implementations. Kenya's approach on comprehensive common dictionary can serve as a model for other countries in similar settings.

Keywords:

Terminology; Electronic Health Records; Concept dictionary; Developing Countries.

Introduction

Developing countries are increasingly adopting information technology applications through the implementation of electronic health records (EHRs) for clinical and managerial activities [1]. This has enabled geographic expansion of access to healthcare, improved data management and reporting, and communication between healthcare providers and their clients [2].

At the heart of each EHR is a concept dictionary, which forms the basis for database organization and semantic interoperability [3]. Concept dictionaries allow the creation of accurate and consistent patient records that can be shared within and across organizations [4]. The use of common terms and concepts in patient records has been shown to enhance the quality of healthcare delivery as numerous decision support systems rely on these terminologies [5].

By their very nature, concept dictionaries are living entities, evolving over time to meet needs within the care organization. Ensuring that dictionaries evolve gracefully usually requires resources and relevant expertise. For a single institution, there are well-defined approaches to assure graceful dictionary management can be achieved [6]. The task however becomes exponentially more complicated when terminologies across multiple institutions have to be semantically interoperable. Dixon et al. succinctly put the problem in context describing the translation of local terminology into available standards as being a complex, costly and resource intensive process [7].

The Unified Medical Language System (UMLS) has allowed the integration of different terminologies without restricting content, structure, or semantics of the original terminologies. This has enabled the creation of mappings among equivalent entities used in different contexts and purposes. This allows semantic interoperability between different terminolgy systems while each evolves to serve its primary purpose [8].

As developing countries start to implement EHRs at scale, they are rapidly running into the issue of assuring semantic interoperability between individual implementations. Unfortunately, clearly defined approaches do not exist to inform how to develop dictionaries to serve multiple implementations within limited resource settings. In most instances, data sharing and system interoperability is only achieved through piecemeal concept mappings. The mappings are often either done between a few local implementations or between a single local implementation's dictionary to multiple standard terminologies [9]. This process is resource intensive and there is often no guarantee that the mappings are accurate [10].

Developing countries with limited resources and few skilled personnel need guidance on approaches that will alleviate the complexity, cost and resource-intensive nature of supporting semantically interoperable dictionaries. In this paper, we describe an approach, taken by Kenya, to come up with a national level concept dictionary to serve multiple EHR implementations across the country. We touch on process, infrastructure, capacity, and foundational issues in creating a national-level concept dictionary that can serve cross cutting needs of clinical care, research, monitoring and evaluation, and reporting. Based on our experience, we outline key principles and approaches that can be used by countries in similar resource-limited settings to comprehensively develop and evolve semantically interoperable dictionaries.

Methods

Setting

Kenya is one of the leading countries in health information technology in Sub-Saharan Africa. Recognizing the important need to manage and use patient data better at the various levels

of care, Kenya's Ministry of Health (MOH) Division of Health Information System developed a Health Information Policy and Strategic plan (2009-2014) [11], and Standards and Guidelines for EHR for the country [12]. After the standardization process was completed, the MOH selected two systems for initial national roll out, namely IQ-Care and KenyaEMR, an adaptation of the open source OpenMRS EHR (http://kenyaemr.org). The MOH also decided that the initial disease foci to be served by these EHR would be HIV, Tuberculosis, and Maternal and Child Health (MCH), with an option to expand to other domains moving forward. Two implementation partners, I-TECH (http://www.go2itech.org/) and Futures Group (http://www.futuresgroup.com/) were selected to implement the selected EHR, and by September 2014, 646 implementations of both systems had been realized at MOH facilities throughout the country.

The national roll out of the two systems occurred in a field that was already increasingly dotted with other EHR implementations. As an example, the Academic Model Providing Access to Healthcare (AMPATH) [13], Kenya Medical Research Institute (KEMRI), and Family AIDS Care and Education Service (FACES), which all offer care services within MOH facilities, already had EHRs with their own local dictionaries.

Challenges

Several challenges related to dictionary management and semantic interoperability quickly became evident in the milieu of multiple EHR implementations. Key challenges included:

- *Inability to share data across EHR:* This example emerged when MOH sites were reorganized to different implementing partners. As an example, it was difficult to easily share data between the Ampath Medical Records System (AMRS) and KenyaEMR systems, despite both being based on the OpenMRS platform due to the different dictionaries in use.
- *Lack of expertise in dictionary management*: With 600 standalone implementations, it was evident that local sites lacked the skill-set to manage dictionaries or new proposed concepts at the local level.
- *Lack of coordination and ownership:* Even when it became clear that a harmonized approach to semantic interoperability was needed across implementations in the country, there were no clear leaders or owners of the problem.
- *Conflict between care, research and reporting needs:* While the MOH aimed to collect data at one reporting level, clinical priorities dictated a higher level of granularity of concepts, while research needs often dictated a broader range of data to be collected beyond what would normally be needed for care or reporting.
- *Lack of policies or processes:* Comprehensive dictionary management requires laid down policies on how to request new concepts, how to make changes to existing dictionaries, and how to prioritize new concept requests for action. In the country, there were big gaps in the relevant policies to guide coordinated management of concept dictionaries.
- *Technical Infrastructure:* The fact that most previously implemented EHRs worked as standalone systems and were not connected to a central server made it difficult to know what the concept needs were at the local level, and also served as a barrier to automatically transmitting concept requests for action centrally.
- *Lack of automated mapping systems:* Even when requests for new concepts could percolate through, their management tended to be largely manual, with deficient mechanisms to automatically map concepts against standardized terminologies.

Approach

Kenya constituted a Concept Dictionary Working group. This working group was made up of 32 individuals with varied but relevant backgrounds, namely: (a) *Healthcare providers* – physicians, clinical officers and nurses; (b) *Data Managers* – Health Records and Health Information Officers, and Monitoring & Evaluation Specialists; (c) *Health Informaticians* - dictionary managers, health system developers and programmers; and (d) *Health Administrators* – MOH County Health Management teams, National EHR implementation coordinators, and program managers.

The Concept Working Group held an in-person meeting between October 29-31, 2014 in Kisumu, Kenya, to define and implement approaches for comprehensive and coordinated national concept dictionary management. This meeting brought together implementing partners and institutions that participate in various healthcare activities as custodians, care providers, information technology solution providers and researchers. Participants were drawn from MOH, I-TECH, AMPATH, KEMRI, FACES, Médecins Sans Frontières (MSF), Elizabeth Glazer Pediatric Aids Foundation (EGPAF), OpenSource Health Management Information System (OpenHMIS) and LakeHUb Kisumu.

The work of the Concept Working Group is largely consultative, but highly informed by well-established lessons, guidance and frameworks around dictionary management.

Results

We describe outcome of the work by Kenya Concept Dictionary initiative, along each of the dimensions outlined in the challenges above.

Common Data Dictionary Considerations

Other than a couple of large institutions in the country, most facilities did not have the skill-set or capacity to create and maintain their own dictionaries. Further, allowing multiple independent dictionaries was already causing interoperability problems. The team recommended that a common national-level concept dictionary be created.

Multiple approaches existed on how to create this common dictionary. Options included:

1. Developing a new dictionary from scratch without influence from existing dictionaries;
2. Taking all existing dictionaries and combining their terms to create a new common dictionary;
3. Using an existing dictionary as the foundation, mapping other dictionaries to it, and evolving it moving forward.

Eventually, the third approach was chosen. This approach is similar to what the Indiana Network for Patient Care (INPC) used when it was coming up with its common dictionary for the Health Information Exchange in the state of Indiana, USA [14]. INPC based its dictionary on the one used by the Regenstrief Medical Record System [15].

It was decided that the Kenya Common Concept Dictionary would be based on the dictionary maintained by Columbia International eHealth Laboratory and the Millennium Villages Project (CIEL/MVP dictionary). The CIEL/MVP dictionary is

already mapped to standard terminology such as SNOMED CT, LOINC, ICD-10, RxNORM and CVX. It also has provisions for Interface Terminology, allowing for easy use within clinical contexts [15]. Given the extensive nature of the CIEL/MVP dictionary, Kenya would only start with those concepts immediately relevant to its needs – these are identified based on clinical needs, relevant MOH forms within the EHR (e.g. MOH 257 – HIV patient care card), previously defined minimum datasets [17], common clinical observations, tests, drugs, procedural observations among others (Figure 1). From this initial subset, some concepts may appear redundant but these will eventually be mapped when future concepts are added as needed. It was also recognized that the CIEL/MVP would not have some terms that would be specific to Kenya, and as such, the Kenyan dictionary would eventually include a small set of its own local terms, but only when these terms were not relevant for incorporation to the larger CIEL/MVP dictionary, and could not be mapped to any other existing vocabulary standard.

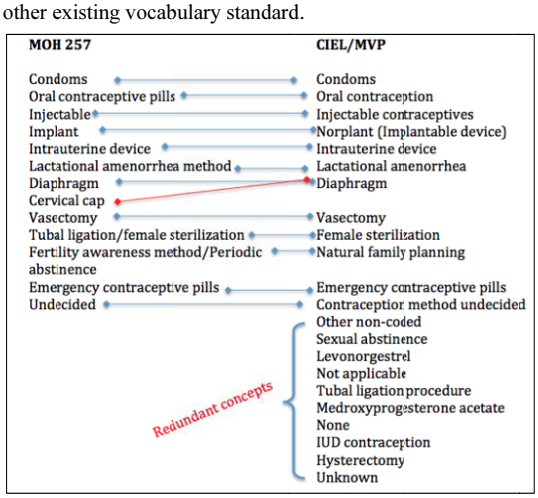


Figure 1: Sample manual mapping of local terms to CIEL/MVP dictionary

Handling conflict between care, research and reporting

A major topic of continued discussion is the scope and granularity level of the Kenya Common Concept Dictionary. There is always tension as to how dictionaries can meet the diverse needs of the various stakeholders, including clinicians, MOH and researchers, among others. Fortunately, Kenya is not the first country faced with this problem. Guidance in this space exist from Chute et al. [18], Cimino's Desiderata [6], and the ISO specifications on terminologies [19]. General recommendations point against creating a single monolithic terminology that attempts to meet all user needs. Instead, terminologies developed should be scoped to serve specific usage categories well [20]. For the case of Kenya, key categories include clinical documentation, administrative, reporting, and research.

Given that EHR being implemented in Kenya are primarily meant to meet clinical needs, a large scope of the Kenya Common Concept Dictionary aims to meet clinical needs. Luckily, key MOH clinical cards (e.g. MOH 257) and registers, which are being replaced by the EHR, already serve the dual purpose of clinical documentation and reporting. We will pay special attention to ensure that we not only have a robust reference terminology, but that it meets interface terminology needs, including appropriate synonyms and multi-lingual functionality. The issues of pre- and post-coordination will continue to be actively considered as the dictionary evolves to meet various needs.

The team recognizes, however, that over time different dictionaries might have to be developed to best serve the various scopes. As an example, a reporting focused dictionary might have concepts that are derivations of multiple concepts in the clinical-targeted dictionary using pre-defined logic. Where appropriate, these dictionaries would also have simple mappings between one other.

Policies and Processes

A core task for the Concept Working Group is to come up with the policies and processes around all aspects of the Kenya Common Concept Dictionary. This organizational aspect of dictionary maintenance is guided by the framework developed by Bakhshi-Raiez et al. [21]. The framework outlines a primary component called 'Execution' which details the various policy and procedural aspects of maintaining a dictionary. Within this framework, criteria and guidelines are provided for submitting new concept proposals, validating them and verifying changes, documentation and version management. In addition to the above, the concept modeling in Kenya is guided by well-established standards for concept creation, curation and evolution [6, 18, 22, 23].

The other components, which support 'Execution', include 'Process Management', 'Change Specifications' and 'Editing Tools'. Under process management, the Concept Working Group has created a team to coordinate its activities, with the custodian institutions being ITECH-Kenya (a leading national EHR implementation) and AMPATH (a large clinical care setting serving a catchment of 3 million people) [24, 25]. Work in progress includes constituting an appropriate maintenance team with procedures to assure responsiveness to requests and creating policies relevant to the dictionary management procedures.

Resources and Human Capacity

Concept dictionary management is resource intensive. The financial support for the Concept Working Group has been provided by the Centers for Disease Control (Grant # U91HA06801). However, extensive resources are still needed to support the executional aspects of this work. Core disciplines that would form the dictionary maintenance team would include user/domain experts, terminology experts, health informaticians, software engineers, coordinators and client support personnel. This formal team has not yet been constituted, and the approach taken so far involves having personnel coming from multiple implementation partners to fulfill these roles. MOH's support and/or custodianship of this initiative would help to assure continued success.

To improve the human capacity to manage dictionaries within the country, special short-courses around Forms, Concepts and Dictionary Management have been conducted by programs such as the Fogarty-funded Regional East African Center for Health Informatics (*REACH-Informatics*) [26]. Emerging programs like the Masters and PhD programs in Health Informatics at Moi University will also have a dedicated coursework that touch on dictionary management - Clinical Decision Support, Ontologies and Workflow (HIC 824). In addition to formal training, the Concept Working Group is leveraging larger informatics networks, especially the OpenMRS community, the OpenHIE Terminology Services [27], the CIEL group, among others to help with various aspects of dictionary maintenance. Implementation partners within the country are also sharing experiences around dictionary management amongst themselves and with the Concept Working Group.

Technical & Infrastructure Consideration

Several technical and infrastructure issues remain relevant to the Kenya Common Concept Dictionary. The dictionary currently does not have a permanent home, with ITECH serving as its custodian. The eventual plan is for the dictionary to be hosted in the National Data Center which is being set up by the Ministry of Health. Software applications still need to be developed to streamline the workflow around concept request, management, mapping and distribution. Mapping, as an example, has remained largely a manual process, with searching done on the Open Concept Lab / Maternal Concept Lab site [28]. In the roadmap for the national concept dictionary is incorporation of semi-automated mapping tools, similar to those employed in other settings and systems [29, 30].

A key technical lesson the country is quickly learning is that EHR implementations that rely on a concept dictionary should not exist without connectivity to a central server. Connectivity is not only essential for health information exchange, but also provides a mechanism for automatically submitting new concept proposals from implementations. Existing approaches to distributing the most current dictionary usually involve emailing of a snapshot, or deployment via SFTP file transfer. Similar to the OpenHIE model, the goal is to eventually be able to deploy the dictionary via API calls from various EHRs.

Discussion

In this paper, we describe efforts by a developing country to implement comprehensive mechanisms for managing concept dictionaries in support of EHR implementations across its MOH facilities. Kenya is not unique in this need. Almost all other developing countries will have similar challenges as they start implementing EHRs at scale. Ideally, countries should anticipate these challenges and needs, and they should put plans for dictionary management in place before glaring challenges emerge. In fact, this issue should be a key focus of initial implementation planning, with countries well advised to appreciate the amount of work and resources required to get it done well. Constituting the right technical working groups, and having appropriate consultations with groups with relevant experience is very important.

Key considerations as outlined in this paper revolve around identifying initial corpus of concepts and assuring that the scope is appropriate for the country's primary needs. Policies and procedures are needed to enable graceful evolution of the dictionary and high quality responsiveness to implementers. Keys to success are the availability of financial resources to support the infrastructure and the right team, recognizing that diverse competencies will be needed. Ideally, the MOH needs to embrace its responsibility as the custodian of this dictionary, and provide the requisite support. This is not to say that the maintenance team has to sit within the MOH, as this responsibility could also be appropriately outsourced to a group that is highly skilled in this area. Capacity building efforts should however be integral to the country's strategy for dictionary management as the responsibilities can require a highly specialized skillset.

Fortunately for developing countries, there is a wealth of resources and research around the optimal approaches for dictionary management for health information exchange semantic interoperability. An appreciation of the core principles, guidelines and frameworks would serve the technical team well [6, 18-20, 22]. Key lessons can also be learned from approaches taken by various Health Information Exchange efforts in the Western world [14]. In addition to consulting developing country partners, countries should also leverage extensive communities around eHealth systems that are working specifically on vocabularies and ontologies. Further, countries are encouraged to share experiences and lessons. Appropriate forums should be constituted to allow these lessons to be shared for multiple levels of stakeholders.

This field is very complex. In fact, there are multiple debates on the optimal approaches to managing dictionaries [31]. The approaches taken by Kenya and presented above might not necessarily be the most optimal in this evolving field. Our team recognizes its limitations, and is willing to adjust direction as needed. The country is also committed to being adaptive and versatile, to assure that it can take advantage of emerging and new approaches to dictionary management. As an example, the OpenHIE Terminology Service is working on a '*Subscription Service*' that would significantly reduce the burden to developing countries in managing concept dictionaries. Kenya's team is already working with this team, and hopes to use this service as needed in managing the Kenya Common Concept Dictionary.

The eventual hope is for countries like Kenya to truly realize truly comprehensive Health Information Exchange systems. The beginning of an Enterprise Health Architecture is starting to be realized in the country, and a comprehensive terminology service would be central to this architecture. In fact, systems like mHealth that are now often implemented in isolation will soon be expected to embrace the same concept dictionary terms that are being used in MOH supported EHR. Obviously, a lot of work remains to help realize the vision of the Kenya Common Concept Dictionary, but the country should be commended for having taken the bold first step in comprehensively addressing the issues around semantic interoperability.

Some of the limitations of this process include the voluntary nature of contributions and participation by individuals and organizations. Required technical and infrastructural supports are not assured. It is hoped that the government can invest more resources and organizations can dedicate more personnel and time in this process.

Conclusion

A Concept Working Group made of multiple stakeholders is leading the evolution of a national level common concept dictionary for Kenya, with close guidance to use well tested approaches for concept dictionary management. The approach by Kenya can be used as a model for other countries hoping to implement terminology based services to support multiple implementations.

Acknowledgments

This work was supported in part by CDC Grant # U91HA06801 and NORAD (Norwegian Agency for Development Cooperation) under the NORHED-Programme , Agreement no. KEN-13/0021. Thanks to the Ministry of Health and Counties for supporting the work of the Concept Working Group and the national EHR rollout. Special thanks go to the following individuals who have been key players in the Concept Working Group: Nicholas Ingosi, Gitahi Ng'ang'a, Kenneth Ochieng, Sara Armson, Petronillah Seurey and Andrew Kanter. We would also like to thank the following organizations for spearheading these efforts and sharing their experiences: I-TECH, KEMRI, FACES, AMPATH and Columbia International eHealth Laboratory and the Millennium Villages Project.

References

[1] Williams F, Boren SA. The role of the electronic medical record (EMR) in care delivery development in developing countries: a systematic review. Inform Prim Care. 2008;16(2):139-45.

[2] Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Organ. 2012;90(5):332-40.

[3] Were MC, Mamlin BW, Tierney WM, Wolfe B, Biondich PG. Concept dictionary creation and maintenance under resource constraints: lessons from the AMPATH Medical Record System. AMIA Annu Symp Proc. 2007:791-5.

[4] Lee MK, Park HA, Min YH, Kim Y, Min HK, Ham SW. Evaluation of the Clinical Data Dictionary (CiDD). Healthc Inform Res. 2010;16(2):82-8.

[5] Monsen K, Honey M, Wilson S. Meaningful use of a standardized terminology to support the electronic health record in new zealand. Appl Clin Inform. 2010;1(4):368-76.

[6] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine. 1998;37(4-5):394-403.

[7] Dixon BE, Vreeman DJ, Grannis SJ. The long road to semantic interoperability in support of public health: experiences from two states. J Biomed Inform. 2014;49:3-8.

[8] Campbell KE, Oliver DE, Shortliffe EH. The Unified Medical Language System: toward a collaborative approach for solving terminologic problems. J Am Med Inform Assoc. 1998;5(1):12-6.

[9] Thibault JC, Frey L. Biomedical Terminology Mapper for UML projects. AMIA Jt Summits Transl Sci Proc. 2013;2013:257-61.

[10] Saitwal H, Qing D, Jones S, Bernstam EV, Chute CG, Johnson TR. Cross-terminology mapping challenges: a demonstration using medication terminological systems. J Biomed Inform. 2012;45(4):613-25.

[11] http://apps.who.int/healthmetrics/library/countries/HMN_KEN_StrPlan_Final_2010_02_en.pdf.

[12] http://www.nascop.or.ke/library/3d/Standards_and_Guidelines_for_EMR_Systems.pdf.

[13] Tierney WM, Rotich JK, Hannan TJ, Siika AM, Biondich PG, Mamlin BW, et al. The AMPATH medical record system: creating, implementing, and sustaining an electronic medical record system to support HIV/AIDS care in western Kenya. Stud Health Technol Inform. 2007;129(Pt 1):372-6.

[14] Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR. Embracing change in a health information exchange. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2008:768-72.

[15] Biondich PG, Grannis SJ. The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. Journal of public health management and practice : JPHMP. 2004;Suppl:S81-6.

[16] Kanter AS, Wang AY, Masarie FE, Naeymi-Rad F, Safran C. Interface terminologies: bridging the gap between theory and reality for Africa. Studies in health technology and informatics. 2008;136:27-32.

[17] Tierney WM, Beck EJ, Gardner RM, Musick B, Shields M, Shiyonga NM, et al. Viewpoint: a pragmatic approach to constructing a minimum data set for care of patients with HIV in developing countries. Journal of the American Medical Informatics Association : JAMIA. 2006;13(3):253-60.

[18] Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-Based Patient Records Institute Working Group on Codes and Structures. Journal of the American Medical Informatics Association : JAMIA. 1998;5(6):503-10.

[19] ISO/TS 17117:2002(E): Health informatics-controlled health terminology-structure and high-level indicators: technical committee ISO/TC 215, Health Informatics; 2002.

[20] Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proceedings / AMIA Annual Symposium AMIA Symposium. 1999:42-6.

[21] Bakhshi-Raiez F, Cornet R, de Keizer NF. Development and application of a framework for maintenance of medical terminological systems. Journal of the American Medical Informatics Association : JAMIA. 2008;15(5):687-700.

[22] Cimino JJ, Clayton PD. Coping with changing controlled vocabularies. Proceedings / the Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care. 1994:135-9.

[23] Oliver DE, Shahar Y. Change management of shared and local versions of health-care terminologies. Methods of information in medicine. 2000;39(4-5):278-90.

[24] Inui TS, Nyandiko WM, Kimaiyo SN, Frankel RM, Muriuki T, Mamlin JJ, et al. AMPATH: living proof that no one has to die from HIV. Journal of general internal medicine. 2007;22(12):1745-50.

[25] International Training and Education Center for Health. http://www.go2itech.org/. Last accessed 12-15-2014.

[26] Regional East African Center for Health Informatics (REACH-Informatics). Available at https://sites.google.com/site/reachinformatics/. Last accessed 12-15-2014.

[27] Open Health Information Exchange (OpenHIE). Available at https://ohie.org/. Last accessed 12-15-2014.

[28] Open Concept Lab. Available at http://openconceptlab.com/. Last accessed 12-15-2014.

[29] Lau LM, Johnson K, Monson K, Lam SH, Huff SM. A method for the automated mapping of laboratory results to LOINC. Proceedings / AMIA Annual Symposium AMIA Symposium. 2000:472-6.

[30] Sun Y. Methods for automated concept mapping between medical databases. Journal of biomedical informatics. 2004;37(3):162-78.

[31] Cimino JJ. In defense of the Desiderata. Journal of biomedical informatics. 2006;39(3):299-306.

Address for correspondence

All correspondence to be addressed to:

Dr Aggrey Keny

AMPATH Centre

P. O. Box 4606

Eldoret, 30100

Kenya. Email: agris57@yahoo.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-785

Evaluation of Herbal and Dietary Supplement Resource Term Coverage

Nivedha Manohar[a], Terrance J. Adam[a,b], Serguei V. Pakhomov[a,b], Genevieve B. Melton[a,c], Rui Zhang[a,c]

[a] *Institute for Heath Informatics, University of Minnesota, MN, USA*
[b] *College of Pharmacy, University of Minnesota, Minneapolis, MN, USA*
[c] *Department of Surgery, University of Minnesota, Minneapolis, MN, USA*

Abstract

The use of Complementary and Alternative Medicine (CAM) is increasingly popular in places like North America and Europe where western medicine is primarily practiced. People are consuming herbal and dietary supplements along with western medications simultaneously. Sometimes, supplements and drugs react with one another via antagonistic or potentiation actions of the drug or supplement resulting in an adverse event. Unfortunately, it is not easy to study drug-supplement interactions without a standard terminology to describe herbal and dietary supplements. This pilot study investigated coverage of supplement databases to one another as well as coverage by the Unified Medical Language System (UMLS) and RxNorm for supplement terms. We found that none of the supplement databases completely covers supplement terms. UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT cover 54%, 40%, 32%, 22% and 14% of supplement concepts, respectively. NDF-RT provides some value for grouping supplements into drug classes. Enhancing our understanding of the gap between the traditional biomedical terminology systems and supplement terms could lead to the development of a comprehensive terminology resources for supplements, and other secondary uses such as better detection and extraction of drug-supplement interactions.

Keywords:

Herbal and nutritional supplements; Complementary and alternative medicine; Biomedical terminology; UMLS; RxNorm.

Introduction

The use of Complementary and Alternative Medicine (CAM) is rising in North America and Europe where conventional western medicine is practiced [1,2]. This is reflected by increasing sales of herbal supplements in the United States and Europe. For example, only 28 percent of women over 60 years in the United States took a Calcium supplement in the early 90s. A decade later, nearly 61 percent of women in that age group were taking Calcium supplements. In the United States, herbal supplement sales have reached an all-time high of 6 billion dollars per year according to a 2013 report published by the American Botanical Council [3]. However, the growing use of supplements is occuring with many adults simultaneously taking prescription medicines. Almost one in six adults in the US taking prescription medicines also takes an herbal supplement [4]. This poses a risk of unintentional interactions between the drug and supplement, especially since documentation is not consistent on supplements [1,3,4]. According to a report published by the Center for Disease Control (CDC), more than half of adults in the U.S take dietary supplements and the use of supplements is on the rise [3]. Fifty-six percent of Europeans are also found to have consumed some form of CAM in 2012 [2]. CAM includes a variety of medical systems, therapies and mind-body interventions. For this paper, we have focused on the subcategory that includes dietary supplements called "biologically based therapies" [1].

Although herbal medicines have gained popularity, knowledge about herbal medicines has not gone up among physicians practicing conventional medicine [4]. In a survey conducted among 1,157 clinicians from around the world, about 75.5 % of them said that doctors are "poorly informed" about herbal medicines. Only 12.9 % said that they always check if their patient is taking herbal medicines before prescribing conventional drugs [5]. There have been some steps to reduce the gap between the use of herbal supplements by consumers and the knowledge of supplements from a clinical perspective.

While herbal supplements are widely believed to be a safe addition to conventional medicines, they can potentially react with western medicine. Herbal and other dietary supplements could increase or antagonize the actions of drugs or they could potentially change pharmacokinetics or pharmacodynamics of the drug or supplement itself [6]. For these reasons, detailed research of drug-supplement interactions is an important area.

In contrast to drug-supplement interactions, drug-drug interactions have been more widely studied and documented. One type of resource leveragd for this, are terminologies which provide a standard and well-defined manner of identifying and describing drugs. For example, RxNorm is a standardized system to describe clinical drugs, developed by the U.S. National Library of Medicine (NLM), which is also the preferred representation of drugs for health information exchange. The brand names, generic names and ingredients are listed along with the relationship between those entities [2,7]. In contrast, there is no well-accepted equivalent for herbal supplements making it more difficult to study drug-supplement interactions. This difficulty creates limitations for studies about drug-supplement interactions and is a substantial barrier for extracting information from the limited existing investigative literature. It is also possible to identify new drug-supplements interactions by extracting information from literature [8] and or collecting through a patient interview [9]. Most studies on drug-supplement interactions focus only on particular known combinations, like cranberry and warfarin, ginkgo and warfarin, or St. John's wort and digoxin [6].

One foundational step towards identifying drug-supplement interactions is establishing a comprehensive terminology system for supplements. This purpose of this paper is to evaluate the supplement term coverage across various online supplements databases. Coverages of the Unified Medical Language System (UMLS) Metathesaurus, Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT), RxNorm and National Drug

File - Reference Terminology (NDF-RT) for representing supplement terminology were also investigated.

UMLS and MetaMap

UMLS is a repository of biomedical vocabularies, which integrates over 2 million names and concepts. It also contains relationships between concepts and brings together many health and biomedical vocabularies and standards. The UMLS includes the NCBI taxonomy, Gene Ontology, MeSH, Online Mendelian Inheritance in Man (OMIM) and SNOMED CT among over 100 other vocabularies [10].

MetaMap is a Natural Language Processing (NLP) application developed by the U.S. NLM to map biomedical texts to the UMLS Metathesaurus concepts [11]. It is able to tag the text lexically and syntactically and identify the different concept candidates in the UMLS. In this study, MetaMap was implemented to map the supplement terms to UMLS concepts.

RxNorm and NDF-RT

RxNorm is a tool developed by the NLM. It has normalized names for branded and generic drug names. Drug names along with strength, brand names, dosage and ingredient information can be found in RxNorm. RxNorm also supports semantic interoperability between drug names and 15 other drug resources proving additional depth of information related to drugs absent with the UMLS [12].

NDF-RT is a part of the Veteran Health Administration's National Drug file. It is able to classify drugs into formal categories in addition to giving information about their molecular interactions, kinetics, therapeutic categories, and dose forms [13].

Supplement databases

Natural Standard Authority Database (NSAD) is a database set up by Natural Standard, an international research collaboration that reviews scientific literature on CAM. This database is arranged alphabetically and includes Spanish and French in addition to English and Latin names. Each entry includes supplement names in other languages, popular brand names, historical significance of the herb, indications for use, expert opinion on efficacy are all provided with citations from peer-reviewed journals [14].

Medscape is a website with medical content from the WebMD Health Professional Network. The Network strives to provide balanced, accurate health information. They provide information written by physicians and other authorities on particular topics making information from Medscape relatively reliable [15].

Natural Medicines Comprehensive Database (NMCD) is a database of nature-based medicines and alternative therapies managed by the therapeutic research center, composed of a panel of pharmacists and clinicians. The panel is not affiliated with any pharmaceutical company, and their work is intended to be useful for pharmacists and clinicians. The website is updated regularly and incorporates the latest research. The database has terms enabling term searches in several Asian and European languages, by botanical names and by common names. Under each plant substance, the effectiveness of the plant, safety, known interactions, mechanism of action, and adverse reactions are listed. There is an application with NMCD that allows users to search for interactions between a drug and a supplement. NMCD is aimed at professional clinicians or pharmacists and may not be easy to understand for a layperson [16].

MedlinePlus is the U.S. National Institute of Health's website for the general public. The website is produced by the NLM and has information about diseases, conditions, health care issues drugs and supplements. The information in the database is available from the U.S. National Center for Complementary and Integrative Health (NCCIH). All statements are supported by evidence from peer-reviewed literature. Supplement names in languages other than English are not available for search. For this study, we only used information from the website on herbs and supplements [17].

Drugs.com is a website that aims to provide consumers with health, drugs and supplement-related information. Information is sourced from Harvard Health publications and Wolters Kluwer Health. Drugs.com provides indications, possible side effects, and interactions with other drugs.It is targeted at lay consumers and provided in a user-friendly format. The website also provides a platform for users to discuss with one another about the drugs. Having only 66 supplements, this was our shortest resource of supplements [18].

Methods

Comparison supplement terms coverages

We compared the supplement term coverage between all five databases in a several step process, as illustrated in Figure 1.

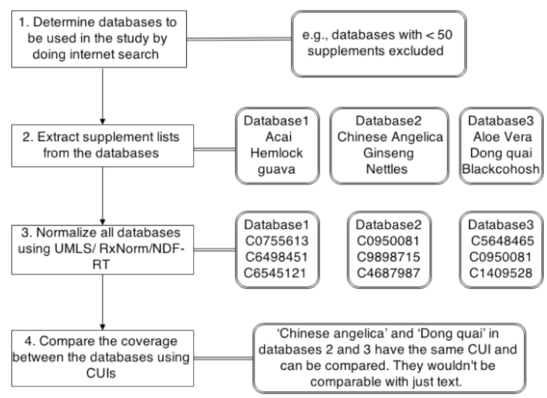


Figure 1 - Illustration of method to compare supplement terms across five online databases

Step 1: Determine the databases to be included in the study. An Internet search was performed for the top herbal supplement databases to identify the included databases. The top results were then subject to further exclusion criteria. If databases were not based on evidence-based science, they were rejected. For example, Holistic heart health, Healthline and WebMD's databases were rejected for not having peer-reviewed sources. Databases with very few (<50) supplements were also rejected. Also one other database on the Dr. Oz's website was no longer available, so that was excluded as well. Rxlist was also rejected since it merely mirrored NMCD's list. Finally, the supplements were obtained from each qualified resource website: NSAD, NMCD, MedlinePlus, Medscape, and drugs.com.

Step 2: Extract supplements from databases. We extracted supplement terms from online databases and used these in our further analysis. We manually removed redundant non-English terms from the list of NSAD while keeping their corresponding English names.

Step 3: *Normalize supplements using UMLS, MeSH and SNOMED CT*. To normalize supplement terms between different databases, supplements were first mapped to UMLS

Concept Unique Identifiers (CUIs) by using MetaMap – which then provides common meanings for the different sources of supplements. Only exact mapping with a score of 1000 were retained. Manual cleanup was performed to remove irrelevant mapping candidates. Semantic types such as 'gene' or 'geographical area' were eliminated before comparing the coverage between different databases. Irrelevant concepts after mapping were also excluded by manual review. When multiple CUIs were returned by MetaMap, the most accurate CUIs were retained. For example, when multiple CUIs were found for a single term, the CUI corresponding to semantic type "plants" was retained over "extracts". Supplement terms were also mapped to MeSH and SNOMED by restricting the terminology source in MetaMap in an effort to use multiple sources to normalize the databases.

Step 4: Compare the coverage between the different databases. Before mapping, we first implemented exact match for supplement terms and analyzed how frequent an individual term exists in each database. UMLS CUIs were then used to compare the coverage of supplements between five databases. We further generate a Venn diagram showing the coverage across databases using UMLS. After integrating concepts from all databases into a full list, we also grouped UMLS concepts based on their semantic types.

Analysis of supplements representation in RxNorm and NDF-RT

To further investigate how RxNorm and NDF-RT represent supplements, we first used RxMix functions to map terms of each database to RxNorm and NDF-RT concepts. RxMix is an NLM developed tool to allow users to combine all functions of RxNorm, RxTerms, and NDF-RT [19]. We then integrated all mapped RxNorm concepts into a respective list. To further investigate the detailed drug class information, we used RxMix to analyze the integrated RxNorm concept list from all databases. Specifically, we created such a workflow to identify the VA drug class for a given supplements string:

$$RxNorm : findRxcuiBySting \rightarrow RxNorm : getRelatedByType \rightarrow NDF-RT : findConceptsByID \rightarrow NDF-RT : getVaClassOfConcept$$

Results

The first part of this study compared supplement terms without referring to any standard terminology source to investigate if any of the databases are able to cover all the supplements from the pooled lists of the databases. Figure 2 shows the frequency in which a particular supplement is found across the 5 databases. Without any term normalization, very few supplements are found across all five databases, but about 2,500 terms appeared in only one of databases.

To compare supplement term coverage across five databases, we normalized terms by using UMLS, MeSH, SNOMED CT, RxNorm, or NDF-RT source. The number of terms mapped to various sources from each database is shown in Table 1. In general, UMLS has the largest coverage for the supplements, and MeSH and SNOMED CT cover more supplement terms than RxNorm and NDF-RT. For example, the largest database NSAD had 1,331 total terms extracted from the website resulting in 1,088 UMLS concepts, 565 RxNorm concepts and 348 NDF-RT concepts with mapping. NMCD had a higher percentage coverage (78%) by UMLS concepts. MedlinePlus and Drugs.com were also covered well by all terminology sources. After integrating supplement terms from different databases into a comprehensive list, we obtained 3,115 unique terms, resulting in 1,683 UMLS, 1,257 MeSH, 984 SNOMED CT, 677 RxNorm, and 422 NDF-RT concepts. Clearly, none of these terminologies can fully represent supplement terms.

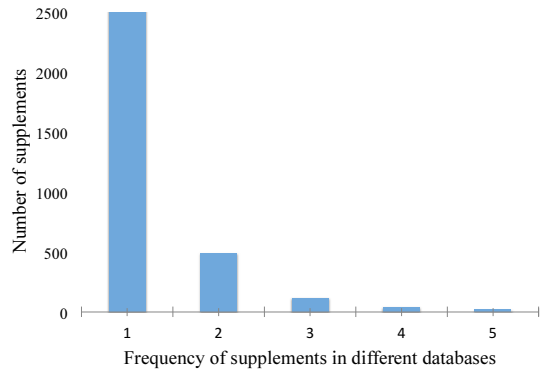


Figure 2 - Frequency of supplement terms across databases

Table 1 – Number of supplement terms mapped to the UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT concepts

	NSAD	NMCD	MedScape	Medline	Drugs
Terms	1331	1038	1080	134	66
UMLS	1088	810	398	121	65
MeSH	930	658	307	108	49
SNOMED	669	492	291	99	41
RxNorm	565	400	228	93	60
NDF-RT	348	272	124	75	43

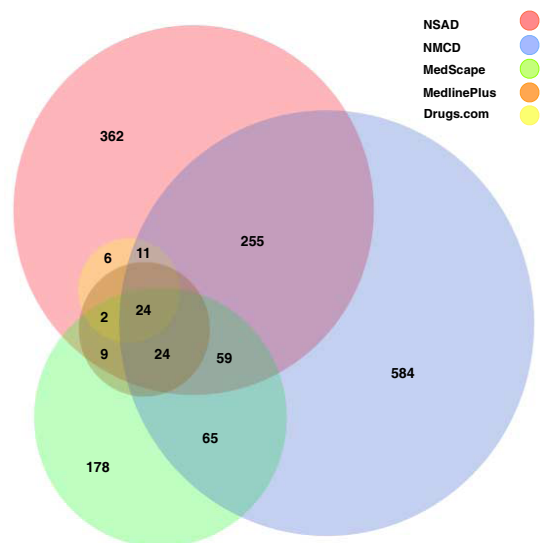


Figure 3 – Venn diagram of UMLS concepts for supplements across five databases

Figure 3 shows a Venn diagram of supplements concepts from the 5 databases after mapping to UMLS concepts. Twenty-four concepts are common to the 5 databases. All five databases have unique CUIs and different coverage with other databases. Not even the largest among the five databases is able to cover a large percentage of terms comprehensively.

After analyzing the semantic types in UMLS concepts, we found that 34% of concepts are plants (*e.g.*, cat's claw, comfrey, damiana), and 18% are organic chemicals (Table 2). Many concepts belong to multiple semantic types, such as Aloe vera, which is a plant, a pharmacologic substance and an organic chemical. About 5% of concepts were classified as food.

Table 2 – Top semantic types of UMLS concepts for supplements

Semantic type	Percentage
Plant	34%
Pharmacologic Substance	27%
Organic Chemical	18%
Food	5%
Biologically Active Substance	4%
Element, Ion, or Isotope	2%
Vitamin	2%

With the aid of the tool RxMix, we found several drug classes for about 300 supplements (about 43% of mapped RxNorm concepts), some of which are listed in Table 3. Not surprisingly, multivitamins were the largest class for supplements. Some other classes, including stimulant laxatives, emollients, antihypoglycemics, antivertigo agents, genito-urinary agents, were also found. Since supplements in the same group can have similar properties, this help to investigate drug functions, adverse effects and interactions with other drugs.

Table 3 – VA drug Class for supplements in NDF-RT

Drug class	Examples
MULTIVITAMINS	alpha-tocopherol,biotin, coenzyme, cyanocobalamin, dexpanthenol, ergocalciferol, folic acid, l-ascorbic acid, niacin, nicotinamide, pantothenol, phytonadione, pteroylglutamic acid, pyridoxine, retinol, riboflavin, thiamine
STIMULANT LAXATIVES	Aloe, cascara, cascara sagrada, castor oil, senna
RESPIRATORY AGENTS,OTHER	Camphor, eucalyptus oil, sodium chloride, turpentine oil
PANTOTHENIC ACID	Dexpanthenol, pantothenic acid, pantothenol, Vitamin B5
EMOLLIENTS	aloe vera, glycerol, retinol, vitamin A, vitamin E
HERBS/ALTERNATIVE THERAPIES	docosahexaenoic acid, eicosapentaenoic acid, ginkgo, ginkgo biloba, inositol, l-methionine, methionine, vitamin E
GENITO-URINARY AGENTS,OTHER	dimethyl sulfoxide, povidone-iodine, yocon, yohimbe, yohimbine

Discussion

This study examines the problem of a lack of standard terminology in describing herbal and dietary supplements. This is a significant problem because herbal supplements may interact with prescription medicines leading to potentially harmful interactions. Without a standardized terminology, studies and comparative research become more difficult, as well as other knowledge-sharing related to supplements.

Although many supplements and their interactions with drugs are published online, they usually use different names and focus on different entities. We found that all databases we examined had substantial gaps. The low degree of supplement terms commonality does create difficulty in utilizing a single database source for supplement study development and evaluation at present. A combined resource approach is likely needed for robust supplement term identification.

Many of the supplement terms could not be mapped to UMLS Metatheasurus. On review of some of the unmapped terms a number of findings were noted. Unlike traditional medications which typically have a generic and trade name and a chemical identifier, the dietary supplements are more complicated especially plant products. Starting with Latin Genus and Species information a particular supplement may have multiple common names depending on where the product is grown and the ultimate supplement consumer. Several common supplements had five of more common names making it challenging to map the name to a single entity. Several exact matchs were also identified as irrelevant mappings. For example, cartilage (bovine and shark) was mapped to "1000 C0007301 Cartilage [Tissue]".

In addition, particular supplements may have different names based on qualifiers for the country of origin or the part of the plant which was used. Some products may have different reference names based on whether it was prepared by some means such as an extract which chemically is similar to the parent supplement but may be changed in some way after preparation. Arguably this may reflect more a dosage change in some cases but will vary on the supplement and the preparation method. Incorporation of the common preparation methods for supplement products may need additional exploration to identify the appropriate standards to accurately reflect the supplement being reviewed.

The number of concepts retrieved by the MetaMap application is much lower than the number of supplements in the NSAD and Medscape databases. One reason is that the MetaMap was also unable to map some chemical names (e.g., 1,3,7-trimethyl-2,6-dioxopurine). Another reason for the observed limited performance of MetaMap is that we restricted matches to exact mapping with a score of 1000. Some concepts with a lower score were therefore excluded. For example, "new zealand green-lipped mussel" was mapped to two UMLS concepts "933 NEW ZEALAND GREEN MUSSEL (New Zealand green mussel extract) [Organic Chemical, Pharmacologic Substance]" and "637 lipped (Lip structure) [Body Part, Organ, or Organ Component]". Another example is a combination of herbs that are available as a single supplement listing – for example, "bilberry/evening primrose/flax", which did not exactly map to any single concept. After manually reviewing those with lower mapping scores (<1000), we found additional 15, 20 and 31 correct mappings for NMCD, NSAD and Medscape, respectively. It was also found that UMLS doesn't have a trinomial classification of plants, making the CUIs less specific. For example, ginseng from different countries (*e.g.*, Chinese gingseng and Korean ginseng) were mapped to the same CUI with outputs as "1000 C0949314: Chinese ginseng (*Panax ginseng*) [Plant]" and "1000 C0949314: Korean ginseng (*Panax ginseng*) [Plant]". Both "ginseng, American" and "ginseng, siberian" were mapped to "C0949314 ginseng (*Panax ginseng*)". In practice, ginseng from different geographical areas is used for different purposes and may have

potentially different active supplement concentrations – arguing for classifying these under different concepts. UMLS perhaps lacks the depth needed to represent herbal supplements. Also, some of the databases used in this study had repetitions containing the same herb in different languages. Place of supplement origin is unique for herbal substances. Such geographic origin information sometimes can be recognized by the supplements names. For example, ma huang has the botanical name *Ephedra sinica* where sinica stands for China. There is another species of Ephedra called *Ephedra funerea* named for the funeral mountains in the United States. Although UMLS had a broader coverage of supplement concepts, RxNorm and NDF-RT, as part of UMLS, can provide more detailed information about medications, including brand and generic name, dosage, drug class and interactions, which is useful for finding relationships between supplements. Future inclusion of additional supplements into these resources and frameworks will have significant value for secondary uses of supplement information. Our study has a number of limitations. One limitation is the use of only five databases, all of which are based in North America. A more comprehensive study would include more databases from geographically diverse sources. While the MetaMap system is accurate, the application is limited to English terms only, and only exact matches were included in this study, resulting in a large number of terms that did not map. The Anatomical Therapeutic Chemical (ATC) classification system [20] devloped by the World Health Organization is another terminology we plan to evaluate the supplements term coverage. Future research will also focus on integrating knowledge resources to form a more comprehensive system of terminology for describing herbal and dietary supplements.

Conclusion

This pilot study compared supplements across five popular online herbal supplement databases. None of the databases were found to be comprehensive enough to represent a large majority overall of supplements in these resources. Analysis of the coverage of UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT for describing herbal and dietary supplements demonstrated that UMLS covered over half (54%) of the supplements extacted and integrated from five databases. RxNorm and NDF-RT contain more detailed drug-related information, including drug classes for some supplements. This study is important because it provides insights on current gaps and potential opportunities to enhance our understanding of how different terminologies can be used to represent herbal supplements.

Acknowledgments

This research was supported by the University of Minnesota Informatics Institute *on the Horizon* grant (RZ), the Agency for Healthcare Research & Quality grant (#1R01HS022085-01) (GM), and University of Minnesota Clinical and Translational Science Award (#8UL1TR000114-02) (Blazer).

References

[1] Ventola CL. Current Issues Regarding Complementary and Alternative Medicine (CAM) in the United States. P&T 2010: 35(8): 461-8.

[2] Zuzak TJ, Boňková J, Careddu D, Garami M, Hadjipanayis A, Jazbec J, et al. Use of complementary and alternative medicine by children in Europe: Published data and expert perspectives. Complement Ther Med 2013: 4:(21) Supplement 1(0): S34-47.

[3] Gahche J, Bailey R, Burt V, Hughes J, Yetley E, Dwyer J, et al. Dietary Supplement Use Among U.S. Adults Has Increased Since NHANES III (1988–1994). April 2011; 2014 (Nov 4, 2014).

[4] Kaufman D, Kelly J, Rosenberg L, Anderson T, and Mitchell A. Recent Patterns of Medication Use in the Ambulatory Adult Population of the United States. The Journal of the American Medical Association 2002: 287(3): 337-44.

[5] http://dtb.bmj.com/site/about/DTB_survey_on_herbal_medicines.pdf.

[6] Gardiner P, Phillips R, and Shaughnessy AF. Herbal and Dietary Supplement–Drug Interactions in Patients with Chronic Illnesses. American Family Physician 2008: 77(1): 73-8.

[7] Bodenreider O, and Peters LB. A graph-based approach to auditing RxNorm. Journal of Biomedical Informatics 2009 April 2004: 42: 558-70.

[8] Zhang R, Adam T, Simon G, Cairelli M, Rindflesch T, Serguei P, et al. Mining biomedical literature to explore interactions between cancer drugs and dietary supplements. AMIA 2015 Joint Summit on Translational science; Mar 2015; San Francisco.

[9] Scarton LA, Zeng Q, Bray BE, and Shane-McWhorter L. Feasibility and potential benefit of collecting complementary and alternative medicine data through a computerized patient interview. *AMIA Annual Symposium Proceedings,* Washington, DC, 2011.

[10] Bodenreider O. The unified medical language system (UMLS): Integrating biomedical terminology. Nucleic Acids Research 2004: 32:267-70.

[11] Aronson AR, Lang F, Aronson AR, and Lang F. An overview of MetaMap: Historical perspective and recent advances. Journal of the American Medical Informatics Association 2010; 17(3): 229-36.

[12] RxNorm. http://www.nlm.nih.gov/research/umls/rxnorm/

[13] NDF-RT. http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/

[14] Natural Standard Authority Database. https://naturalmedicines.therapeuticresearch.com/

[15] MedScape. http://www.medscape.com/public/about

[16] Natural Medicines Comprehensive Database. http://naturaldatabase.therapeuticresearch.com/

[17] MedlinPlus. http://www.nlm.nih.gov/medlineplus/druginformation.html

[18] Drugs.com. http://www.drugs.com/support/about.html

[19] RxMix. http://mor.nlm.nih.gov/RxMix/

[20] ATC. http://www.whocc.no/atc_ddd_index/

Address for correspondence

Rui Zhang, PhD, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN, USA, 55455. Email: zhan1386@umn.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-790

Semantic Alignment between ICD-11 and SNOMED CT

Jean-Marie Rodrigues[a,b], David Robinson[c], Vincenzo Della Mea[d], James Campbell[e], Alan Rector[f], Stefan Schulz[g], Hazel Brear[h], Bedirhan Üstün[i] , Kent Spackman[c] , Christopher G. Chute[j] , Jane Millar[c], Harold Solbrig[k], Kristina Brand Persson[l]

[a] *INSERM U1142, LIMICS, Paris, France*
[b] *Department of Public Health and Medical Informatics, Univ. Jean Monnet of Saint Etienne, France*
[c] *Intl Health Terminology Standards Development Organization, Copenhagen Denmark*
[d] *Department of Mathematics and Computer Science, University of Udine, Italy*
[e] *Department of Internal Medicine University of Nebraska Medical Center Omaha USA*
[f] *University of Manchester, United Kingdom*
[g] *IMI, Medical University of Graz, Austria*
[h] *Health and Social Care Information Centre. United Kingdom*
[i] *World Health Organization, Geneva, Switzerland*
[j] *Division of General Internal Medicine, Johns Hopkins University, Baltimore, MD, 21287, USA*
[k] *Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA*
[l] *National Board of Health and Welfare, Stockholm, Sweden*

Abstract

Due to fundamental differences in design and editorial policies, semantic interoperability between two de facto standard terminologies in the healthcare domain – the International Classification of Diseases (ICD) and SNOMED CT (SCT), requires combining two different approaches: (i) axiom-based, which states logically what is universally true, using an ontology language such as OWL; (ii) rule-based, expressed as queries on the axiom-based knowledge. We present the ICD–SCT harmonization process including: a) a new architecture for ICD-11, b) a protocol for the semantic alignment of ICD and SCT, and c) preliminary results of the alignment applied to more than half the domain currently covered by the draft ICD-11.

Keywords:

ICD, SNOMED CT, Standards, Ontology, Terminology, Classification.

Introduction

The project to achieve semantic alignment between these two standards in the healthcare clinical vocabulary began with an agreement signed in 2010 between the World Health Organization (WHO) and the International Health Terminology Standards Development Organization (IHTSDO). ICD[1], currently published as ICD-10, is the most important worldwide standard for mortality and morbidity statistics. However, it is is also used – in several national modifications and extensions – for health care documentation and billing. The international clinical terminology standard SCT[2,3] has been expanding under the management of the IHTSDO. SCT promises to provide an international standard for codes, terms and formalisms to represent details of the health care process.

The current ICD – SCT alignment efforts occur at a time when clinicians, documentation specialists, epidemiologists, health care administrators and health service researchers identify more and more use cases in which SCT is used in parallel with ICD and local procedure and medication terminology systems. This alignment is driven by requirements for increasing granularity of clinical content to record expanding medical knowledge arising from genomic and related research. To ensure full semantic interoperability between ICD and SCT, a semantic alignment policy was developed which relates ICD classes to rule-based queries depending upon an ICD-11–SCT Common Ontology (CO) [4]. Here we report on the current state of this harmonization effort.

This harmonization requires an innovative architecture for ICD-11 because, in the past, the two standards have been based on different semantics: SCT on axioms that express universal truths (e.g. that all instances of *Thrombosis* affect the vascular system); ICD on rule-based knowledge that introduce class definition (e.g. thrombosis in pregnancy falls into a different class for public health reporting).

Materials and Methods

ICD-11 – SCT Harmonization

In 2007, the WHO launched the revision of ICD[5]. After the agreement between WHO and IHTSDO, a Joint Advisory Group (JAG) was established in 2010. There was consensus within JAG that the harmonization could not simply be a mapping between representational entities (classes and concepts) of both systems. The consensus approach was to base the alignment around a Common Ontology following widely acknowledged principles [6-10].

ICD-11 was designed as a multi-component architecture[4]. The first component is a set of "linearizations" for different uses cases – mortality, morbidity, primary care – that are organised as a single hierarchy with disjoint, exhaustive classes taking origin in previous versions of ICD. A second component, named the Foundation Component (FC), contains all of the ICD-11 classes organised according to new, more flexible principles.

This foundation component has at its core, a model of meaning based on description logic [11], using formalisms and language equivalent to those of the semantic web community deployed in OWL[12] and SNOMED CT. This model was named the ICD – SCT Common Ontology (CO)[4, 13]. Fig. 1 illustrates how the common ontology is related to: (i) SNOMED CT, (ii) ICD-11 linearizations and (iii) contingent

knowledge in the ICD content model, such as diagnostic criteria or therapies, originating with WHO class definitions.

The Common Ontology is a subset of the international release of SNOMED CT (hereafter abbreviated to "SNOMED") expanded and revised for ICD convergence. The Common Ontology has been harmonized with ICD text definitions supplied by the WHO. The CO drew primarily from SNOMED *Clinical Findings* hierarchy, which includes findings, disorders and diseases. The CO has minor components from other SNOMED hierarchies including *Situations*, *Events* and *Social context* and will have defining attributes taken from *Body Structure, Organisms, Physical agents* and others.

JAG had concluded during convergence discussions that these concepts denote clinical situations, i.e. phases of a patient's life, in which a given condition of clinical relevance is present [14].

The ICD class definitions and metadata were assembled using the ICD URI API[15, 16]. SNOMED normal forms and definitions were provided by IHTSDO from the 2015-01-30 international release. The two terminologies were lexically mapped and managed with an Equivalence Table (ET), which was the worksheet for semantic analysis as described below. A Sequel Pro API was used by the IHTSDO to interface with a DL classified developmental version of SNOMED.

This ET contained stated normal forms of pre-coordinated concepts as well as proposed additions to SNOMED. Referential quality assurance rules ensured consistency between chapters and tracked changes to SNOMED across developmental releases.

The architecture of the system was built around a web-accessible MySQL ET data base that could be fed with Excel files or SQL. The database could generate output in any of these modes or as an OWL file [17]. The database is synchronised with IHTSDO equivalence matching tools using a customized exchange format. In this database, we used a double browser (ICD-11/SNOMED) with graphical interface connected by equivalences links.

The web application was able to maintain multiple equivalences, recorded by author, in order to also study inter-observer agreement in equivalence identification. When testing semantic alignment required reclassification of the common ontology, we exported an OWL version to Protégé [18] for description logic classification and comparison of inheritance.

Methods for semantic alignment

1 For a defined subset of ICD beta foundation hierarchy (roughly equivalent to a chapter in ICD-10), generate a candidate map from ICD-11 classes to concepts in "Clinical findings", "Situations", "Events" or "Social context" branches of the SNOMED hierarchy. To identify the map, consider the SNOMED fully specified name (FSN), ICD short text definition, the SNOMED logical definition, and the SNOMED Short Normal Form. (Class M, Table1)

2 For ICD-11 classes without corresponding SNOMED content, mark as Unmatched (U). Develop when possible a candidate pre-coordinated SNOMED concept node to be added to core. Use the new SNOMED concept's normal form as the Common Ontology (CO) concept. (class U/A see Table 1)

3 If ICD class is too complex for a single or pre-coordinated SNOMED concept, try to express the ICD 11 class as a Boolean Logical expression within the constraints of SNOMED model of meaning. Identify the expression as the CO entry. (class U/E, Table 1)

4 Bypass ICD-11 residual classes (NEC) but check if there is a broader match (Parent) (Class U/R, Table 1)

5 If none of the above is possible, propose added SNOMED attributes (U/X, Table 1) or new attribute values (U/EX, Table 1) to create the CO concept.

Table 1– Types of match of ICD Common Ontology concepts to SNOMED CT (SCT)

Match Type & Meaning	**Action in SNOMED**	**Common Ontology Axiom**
Match (M)	–	SCT Short Normal Form
Unmatched/A (U/A)	Add appropriate pre-coordinated concept to SCT	New SCT precoordinated Normal Form
Unmatched/E (U/E)	Post-coordinated expression without change in the model of meaning	SCT post-coordinated Logical expressions
Unmatched/R (U/R)		None
Unmatched/X (U/X)	Potential to add with change to SCT model of meaning	Discussion with IHTSDO
Unmatched /EX	Potential to add with change to content model-object/value	Discussion with IHTSDO

Table 1 summarizes the different types of SNOMED (SCT) candidate matches to Common Ontology

6 For each pair of ICD-11 class/subclass and SNOMED concept in the equivalence table: a) check the WHO short text definition for content, consistency and meaning; b) check the semantics of the SNOMED concept or expression including FSN and description logic (DL) definition (short normal form) to assess the alignment of the meanings of the ICD and SNOMED definitions; c) flag all discrepancies and send them to the WHO/IHTSDO interdisciplinary team for:

a) modification of ICD-11 text definition by a Joint Advisory Group definitions workgroup, or

b) changes to SNOMED description logic definition by SNOMED editors

7 For revisions to SNOMED concept definitions, recompile the DL classification of the edited SNOMED content including expressions (U/E). From the reclassified SNOMED, enumerate the set of all subsumed SNOMED concepts corresponding to each equivalent ICD-11 class and assure that the subsumed set has a one-to-one match within the set of subsumed SNOMED mapped concepts and expressions. Identify discrepancies between the subsumed sets.

8 Evaluate the discrepant class/concept pairs for FSN and logic definitions and determine the root cause of the mismatch. Is this is a misalignment of the ICD-11 subclass with the definition of the ICD-11 class or a difference in concept or attribute definition in SNOMED?

Submit the discrepancies to an interdisciplinary team to identify the root cause and propose:

a) JAG workgroup to revise ICD-11 class definition

b) IHTSDO Editorial staff to revise or expand SNOMED content

9 After resolution of all issues in this chapter of ICD-11, return to step 1 for analysis of next domain.

Results

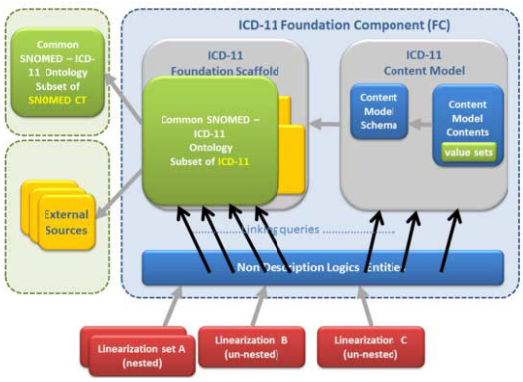


Figure 1 – Architecture of the new ICD-11 architecture and its relationship to SNOMED CT

Updated architecture

Figure 1 is a graphical representation of the new ICD-11 architecture. The outer light blue box represents the Foundation Component (FC), which contains all of the ICD-11 classes or entities, regardless of source. The two smaller boxes on the left represent the external sources for concepts, primarily SNOMED. The Common Ontology is shown in green, a subset of SNOMED participating in ICD-11.

Linearizations are shown in red below the Foundation Component. Each code in each linearization points to a "non-DL entity" in the Foundation Component (FC) whose meaning in terms of the Ontology Component is expressed in a linking query referencing the Ontology Component.

Additional facts about the entities in the Foundation Component are contained in the Content Model. These might include signs and symptoms of diseases, indications and contraindications of drugs among many. These facts are "contingent" (or in traditional philosophical language "particular") rather than universal. That is, they are usually true, or true under certain conditions, but not necessarily "true by definition".

Semantic alignment

The semantic alignment process is ongoing; to date: 16 751 classes/subclasses of the ICD-11 Foundation component have been studied from approximately 39,000 entities (42,9%) .

Table 2 summarizes the percentage of each ICD-11 chapter that has been analysed.

Tables 3, 4 and 5 provide examples of semantic alignment activities to date.

Table 6 presents the summary results of match types for the portion of ICD-11 evaluated thus far.

Table 2– Semantic alignment by ICD-11 chapter (percentage on 01-01-2015)

Chapter	(%)
Infectious diseases	90%
Diseases of the blood and blood-forming organs	30%
Endocrine, nutritional and metabolic diseases	55%
Diseases of the ear and mastoid process	60%
Diseases of the circulatory system	100%
Diseases of the respiratory system	100%
Diseases of the digestive system	100%
Diseases of the skin	80%
Diseases of the musculoskeletal system and connective tissue	90%
Diseases of the genitourinary system	100%
Pregnancy, childbirth and the puerperium	100%
Certain conditions originating in the perinatal and neonatal period	100%

Table 3– Exact match examples

ICD11 Rubric	Common Ontology (Short Normal Form)
Cerebral venous thrombosis	95464572001|Disease (disorder)|:{116676008|Associated morphology (attribute)|=396339007|Thrombus (morphologic abnormality)|363698007|Finding site (attribute)|=68351006|Structure of cerebral vein (body structure)|}
Coronary vaso spastic disease with angina	194828000|Angina (disorder)|+23687008|Coronary artery spasm (disorder)|:{363698007|Finding site (attribute)|=74281007|Myocardium structure (body structure)|}

Discussion

The overall results show that among 16,751 ICD-11 FC entities 14,348 (85.5%) (Table 6) can be represented by the SCT model of meaning either directly or through a pre-coordinated/post–coordinated alignment between the two systems. Very few are residuals which have to be cleaned from the FC or need a revision of SCT formal model of meaning. We propose three examples (Tables 3, 4 and 5) to show how the two systems share the same universal knowledge and how they differ in their contingent knowledge:

1. The defining relationships of SNOMED allow the full representation that can be used for both the ICD-11 Foundation Component class and the SNOMED concept. For example, the ICD Foundation Component entity, "Coronary vaso-spastic disease with angina", is necessarily and sufficiently defined by SNOMED "87343002 | Prinzmetal angina (disorder)".
2. The SNOMED defining relationships do not provide a complete ontological representation of both entities in a single SNOMED concept, in which case the concept model or pre-coordinated SNOMED content must be

Table 4– Pre coordination examples

ICD-11 Rubric	Common Ontology
Aldosterone-producing carcinoma	FSN Primary hyperaldosteronism due to aldosterone-secreting malignant neoplasm of adrenal gland (disorder) Short Normal Form 116680003 | Is a |88213004 | Hyperaldosteronism, |42752001 | Due to|255035007 | Adrenal carcinoma|
Acute myocardial infarction, STEMI, anterior wall	FSN Acute ST segment elevation myocardial infarction of anterior wall(disorder) Short Normal Form 401303003 | Acute ST segment elevation myocardial infarction | + 54329005 | Acute anterior myocardial infarction

Table 5– Post coordination examples

ICD-11 Rubric	Common Ontology
Asymptomatic stenosis of extracranial carotid artery	116680003 | Is a |230738008 | Asymptomatic cerebrovascular disease, |363698007 | Finding site|17999001 | Structure of cervical portion of internal carotid artery, |116676008 | Associated morphology|415582006 | Stenosis
Internal auditory artery occlusion	116680003 | Is a |2929001 | Occlusion of artery, |363698007 | Finding site|89471000 | Structure of labyrinthine artery

expanded. An Example is ICD-11 "Acute myocardial infarction, STEMI anterior wall" that can only be represented in SNOMED by pre-coordinating "401303003 |Acute ST segment elevation myocardial infarction (disorder)" with "54329005 | Acute anterior myocardial infarction".

Conclusion

The essence of the ICD-11 SCT semantic alignment is the establishment of a SNOMED subset with its logical or model of meaning representation that precisely formalizes the meaning of the content of the ICD-11 Foundation Component, following principles of formal ontology and logic *i.e.* that is restricted to axioms that express universal truths in terms of SNOMED concepts. This is clearly distinguished from the ICD content model on the one hand, which represents contingent knowledge at the level of Foundation Component entities, and the rules base (Fig. 1, "non-DL entities"), that cannot be expressed directly in the SNOMED compositional grammar (or any similar logical formalism) and which contains queries on the common ontology that assure the disjointness principle in the linearizations created out of them.

Thus, all content of ICD-11, the semantic standard for health statistics in mortality, morbidity, primary care documentation and billing, will be linked to SCT, the most fine grained medical terminology system, each of which keeps its own profile as a distinct terminology artifact.

Table 6– Match types overall results

Match Type	Number	Common Ontology
Match M	8354 (49.8%)	SCT Short Normal Form
U/A	4933 (29.4%)	To be developped with SCT grammar and pre-coordination
U/E	1061 (6.3%)	To be developped with SCT grammar and post coordination
U/R	1487 (8.8%)	Navigational/residual concepts
U/X and U/EX	916 (5.4%)	Requires clarification

This will require certain refinement and redesign efforts increasing the quality on both ICD-11 and SNOMED, but this is an advantage in itself. When finished, users will have at their disposal two semantically interoperable terminology systems, each tuned for its specific purposes. In the longer term, sharing the maintenance between WHO and the SNOMED authority, IHTSDO, will ease the introduction of new knowledge sources into the heathcare community.

Further on, this common ontology shall be used for the maintenance of all of the existing WHO ICD as well as the ICD-(10/11) national modifications, thereby easing international comparisons and backward compatibility with current systems.

Acknowledgments

This work was supported by the World Health Organization (WHO) and the International Health Terminology Standards Development Organization (IHTSDO) through their Joint Advisory Group (JAG).

References

[1] The World Health Organization, The International Classification of Diseases 11th Revision is due by 2017, http://www.who.int/classifications/icd/revision/en/ (last accessed 15 Dec 2014).

[2] International Health Terminology Standards Development Organisation (IHTSDO): SNOMED CT: http://www.ihtsdo.org/snomed-ct/ (last accessed 16 Dec 2014).

[3] SNOMED CT browser: https:/uts.nlm.nih.gov/ (last accessed 15 Dec 2014)

[4] J.M.Rodrigues, S.Schulz, A.Rector, K.Spackman, T.B.Üstün, C.G.Chute, V.Della Mea, J.Millar, K.Brand Persson, Sharing Ontology between ICD-11 and SNOMED CT will enable seamless re-use and semantic interoperability. Studies in Health Technology and Informatics 2013;192:343-346.

[5] J.M.Rodrigues, A.Kumar, C.Bousquet, B.Trombert, Using the CEN/ISO standard for categorial structure to harmonise the development of WHO international terminologies, Studies in Health Technology and Informatics 150 (2009), 255-259.

[6] B.Smith, Applied Ontology: A New Discipline is Born, Philosophy Today, 29(12) (1998), 5-6.

[7] M.Ashburner et al. Gene Ontology: tool for the unification of biology. Nature Genetics 25(2000), 25:29.

[8] OBO foundry principles: http://obofoundry.org/ontologies.shtml (last accessed 5 Oct 2014).

[9] A.Rector, A.Rossi, M.F. Consorti, P.Zanstra, Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation, International Journal of Medical Informatics 48 (1998), 71-84.

[10]S.Schulz, R.Cornet, K.Spackman, Consolidating SNOMED CT's ontological commitment, Applied Ontology 6 (2011), 1-11.

[11]F.Baader et al. The Description Logic Handbook: Theory, Implementation, and Applications, 2nd Edition. Cambridge University Press, 2nd ed., 2007.

[12]W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview 2009, http://www.w3.org/TR/owl2-overview/ (last accessed 28 Oct 2014).

[13]Rodrigues, J. M., Schulz, S., Rector, A., Spackman, K., Millar, J., Campbell, J., & Persson, K. B. (2013). ICD-11 and SNOMED CT Common Ontology: Circulatory System. Studies in health technology and informatics, 205, 1043-1047.

[14]S.Schulz, A.Rector, J.M.Rodrigues, K.Spackman, Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012, 819-827.

[15]The World Health Organization, ICD-11 beta draft: http://www.who.int/classifications/icd11 (last accessed 28 Apr 2014).

[16]G.Jiang, H. Solbrig, C.G. Chute, "Building Standardized Semantic Web RESTful Services to Support ICD-11 Revision", Proc of SWAT4LS, 2012, CEUR-WS 952.

[17]V.Della Mea, O.Vuattolo, B.Ustun ,A.Rector, S.Schulz, J.M.Rodrigues,K.Spackman, D.Robinson, J.Millar, J.R.Campbell,C.G.Chute, H.Solbrig, K.B.Persson, A web-based tool for supporting the development of a Common Ontology between ICD11 and SNOMED-CT. Proceedings of 2nd IEEE International Conference on Healthcare Infomatics (ICHI2014), 144-148, IEEE, 2014

Address for correspondence

RODRIGUES Jean-Marie <rodrigues@univ-st-etienne.fr>

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-795

Extending the coverage of phenotypes in SNOMED CT through post-coordination

Ferdinand Dhombres[a], Rainer Winnenburg[b], James T. Case[a] and Olivier Bodenreider[a]

[a] *Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, USA*
[b] *Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA*

Abstract

***Objectives**: To extend the coverage of phenotypes in SNOMED CT through post-coordination. **Methods**: We identify frequent modifiers in terms from the Human Phenotype Ontology (HPO), which we associate with templates for post-coordinated expressions in SNOMED CT. **Results**: We identified 176 modifiers, created 12 templates, and generated 1,617 post-coordinated expressions. **Conclusions**: Through this novel approach, we can increase the current number of mappings by 50%.*

Keywords:
phenotype; SNOMED CT; ontology; post coordination

Introduction

While the usefulness of coarse phenotyping based on electronic health record (EHR) data has been demonstrated in the context of recent genomic studies (e.g., [1]), the study of rare syndromes requires detailed phenotyping. More generally, deep phenotyping is required in order to understand how genetic variation relates to clinical manifestations [2]. Despite efforts to facilitate the adoption of standards for phenotyping across domains (e.g., PhenX project [3]), resources for phenotyping tend to vary between clinical data repositories used for translational research and in healthcare settings. For example, while somewhat overlapping, the Human Phenotype Ontology (HPO) used for annotation of research data and SNOMED CT used in EHRs have not been developed in a coordinated fashion and are only partially interoperable.

In previous work, we assessed the coverage of HPO terms in standard terminologies using simple lexical mapping through the UMLS [4]. Only 54% of HPO classes mapped to UMLS concepts and only 30% mapped to SNOMED CT. This simple approach only considered mapping to pre-coordinated terms in SNOMED CT. However, in addition to the pre-coordinated terms distributed with the terminology, SNOMED CT supports the creation of post-coordinated expressions, i.e., logical definitions based on the SNOMED CT concept model. The examination of HPO terms with no lexical mapping to SNOMED CT reveals that they can often be decomposed into simple elements, which could be mapped to SNOMED CT and aggregated into post-coordinated expressions.

For example, the HPO term "Renal Hypoplasia" [HPO:HP_0000089] maps to the (pre-coordinated) SNOMED CT concept "Congenital hypoplasia of kidney" [SCTID:32659003]), with synonym "Renal Hypoplasia". In contrast, there is no pre-coordinated concept in SNOMED CT for the HPO term "Macular hypoplasia" [HPO:HP_0001104].

However, as shown with "Congenital hypoplasia of kidney", the notion of congenital hypoplasia can be represented in SNOMED CT, which also provides a concept for the anatomical structure "Macula lutea structure" [SCTID:82859000]. Therefore, it is possible to create a post-coordinated expression for "Macular hypoplasia", using the template provided by "Congenital hypoplasia of kidney".

The main objective of this work is to extend the coverage of phenotypes in SNOMED CT through post-coordination, i.e., beyond simple lexical mapping to pre-coordinated terms. More specifically, we remove various types of modifiers in HPO terms in order to decompose them into semantic elements that can be recomposed into SNOMED CT expressions through post-coordination. We demonstrate that we can increase the current number of mappings by 50%.

Background

HPO. The Human Phenotype Ontology (HPO) is an ontology of phenotypic abnormalities, used for the annotation of databases such as OMIM (Online Mendelian inheritance in Man), Orphanet (knowledge base about rare diseases), and DECIPHER (RNAi screening project) [5]. The version of HPO used in this investigation is the (stable) OWL version downloaded on April 16, 2014 from the HPO website. It contains 10,491 classes and 16,414 names for phenotypes, including 5,923 exact synonyms, in addition to one preferred term for each class.

SNOMED CT. Developed by the International Health Terminology Standard Development Organization (IHTSDO), SNOMED CT is the world's largest clinical terminology and provides broad coverage of clinical medicine, including findings, diseases, and procedures for use in electronic medical records [6]. It is implemented with a description logic backend and supports two types of concepts. Pre-coordinated concepts are named and defined in SNOMED CT. Post-coordinated concepts are made up of other concepts in a compositional approach [7]. The U.S. edition of SNOMED CT dated September 2013 is used in this work.

UMLS. The Unified Medical Language System (UMLS) is a terminology integration system developed by the U.S. National Library of Medicine [8]. The UMLS Metathesaurus integrates many standard biomedical terminologies, including SNOMED CT. Although the UMLS does not currently integrate HPO, it is expected to provide a reasonable coverage of phenotypes through its source vocabularies. In the UMLS Metathesaurus, synonymous terms from various sources are assigned the same concept unique identifier, creating a mapping among these source vocabularies. Terminology services provided by the UMLS support the lexical mapping of terms to UMLS concepts. Additionally, each UMLS concept is assigned one of the 15 Semantic Groups, which represent broad domains, including Disorders, Anatomy and Physiology. The 2013AB version of the UMLS is used in this work.

Related work

Researchers have investigated the representation of phenotypes through pre- and post-coordinated terms. Groza et al. developed an automated approach for decomposing skeletal phenotype concepts defined in HPO [9]. Oellrich et al. proposed an automated transformation of pre-coordinated phenotypes into Entity-Quality (EQ) statements for achieving interoperability between phenotype ontologies and compare their results to manually created EQ statements available for about half of the terms in HPO [10].

Besides our previous study [4], there have been a few efforts to study the coverage of phenotypes in standard terminologies such as SNOMED CT. Sollie et al. found that there are sizable gaps in SNOMED CT for metabolic disorders (and possibly for other classes of rare and genetic disorders) [11]. However, when Beck et al. assessed the suitability of HPO, ICD10, MeSH, SNOMED CT, and the Human Disease Ontology (DO) for describing GWAS traits, they concluded that, despite generally poor coverage (~20%), partial term matching to SNOMED CT is most successful [12].

These findings suggest, that there is a need for systematic improvement of the coverage of phenotypes in SNOMED CT, and that such an effort might benefit from decomposition techniques similar to prior work [9,10].

Specific contribution

The specific contribution of this work is to identify templates in SNOMED CT for the creation of post-coordinated expressions for phenotype concepts from HPO. This approach significantly extends the mapping to pre-coordinated concepts used in most mapping studies, including our earlier work on the coverage of HPO terms in SNOMED CT.

Methods

Our approach to assessing the coverage of HPO phenotypes in SNOMED CT can be summarized as follows. Starting from HPO terms with no lexical mapping to pre-coordinated SNOMED CT, we identify frequent modifiers and transformation rules in order to make HPO terms compatible with SNOMED CT. We examine the logical definition of existing phenotype concepts in SNOMED CT for templates. We then apply the transformation rules in order to decompose the HPO terms into simple elements. Finally, we combine them again into post-coordinated expressions consistent with the templates.

Establishing a list of modifiers and transformation rules for HPO terms

Starting from HPO terms with no lexical mapping to pre-coordinated SNOMED CT through the UMLS, we analyze word frequencies in order to identify frequent modifiers. More specifically, one author (FD) manually reviewed samples of terms containing words occurring more than 25 times in HPO terms (150 words). Frequent words include “abnormality [of]” (1126), “aplasia/hypoplasia” (134) and “congenital” (44). In addition, we noted frequent abbreviations for ordinals (e.g., “2nd”), while SNOMED CT typically uses the extended form (“second”).

We create four types of lexical transformations for making HPO terms compatible with SNOMED CT. These transformations are presented in increasing order of aggressiveness.

Level 1: Replace

We simply replace the abbreviations for ordinals by their expanded form. For example, the HPO term “Aplasia of the phalanges of the 4th toe” is transformed into “Aplasia of the phalanges of the fourth toe”.

Level 2: Split

HPO uses “/” for coordinating variants (e.g., “aplasia/hypoplasia”), while SNOMED CT does not. Here, we interpret “/” as disjunction, i.e., “aplasia or hypoplasia”. Therefore, we split expressions with words concatenated by “/” into two individual words, which we substitute to the corresponding expression in the term. For example, we transform the term “aplasia/hypoplasia of the thymus” into two terms, “aplasia of the thymus” and “hypoplasia of the thymus”, which both map to SNOMED CT.

Level 3: Demodification to disorder (D0)

HPO contains modifiers that specialize disorders (e.g., “congenital” specializes “bilateral cataract” to form “bilateral congenital cataract”). By removing these modifiers, we create a more general concept, which is more likely to map to SNOMED CT. For example, “bilateral cataract” maps to SNOMED CT.

Level 4: Demodification to anatomy, physiology or chemical substance (D1)

Another category of modifiers are those that denote a disorder (e.g., an abnormality) of a specific anatomical structure, physiological process or chemical substance. Examples of such terms include “Abnormality of the lip”, “Abnormality of intracranial pressure” and “Abnormality of prothrombin”. By removing “abnormality [of the]”, we extract the anatomical structure (e.g., lip), physiologic process (e.g., intracranial pressure) or chemical substance (e.g., prothrombin), which is the object of the abnormality.

Identifying templates for phenotype concepts

As shown in the example presented in the introduction, some pre-coordinated SNOMED CT concepts include the very modifiers (or constructs) we have identified as generally preventing the mapping to SNOMED CT. For example, the SNOMED CT concept “Congenital hypoplasia of kidney” contains “Hypoplasia”, which we have identified as a modifier (at the ***D1*** level). Such SNOMED CT concepts therefore suggest valuable templates for the decomposition of similar concepts. For example, the logical definition of the concept “Congenital hypoplasia of kidney” in SNOMED CT suggests a template for the modifier “Hypoplasia”.

***'Disease (disorder)'** and ('Role group (attribute)' some (('Associated morphology (attribute)' some **'Hypoplasia (morphologic abnormality)'**) and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some **'Kidney structure (body structure)'**)))*

The template suggested by this SNOMED CT concept is “Congenital hypoplasia of <ANATOMICAL STRUCTURE>” and corresponds to the following definition:

***'Disease (disorder)'** and ('Role group (attribute)' some (('Associated morphology (attribute)' some **'Hypoplasia (morphologic abnormality)'**) and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some **'<ANATOMICAL STRUCTURE> '**)))*

where <ANATOMICAL STRUCTURE> is what remains from the original term after demodification.

Applying transformation rules to HPO terms

We apply the transformation rules to HPO terms in increasing order of aggressiveness, from level 1 to level 4. Moreover, at a given level, the transformation rule is applied to not only the original terms, but also to the terms produced by all previously applied rules (at lower levels).

In practice, as shown in Figure 1, we first apply the ***Replace*** rule to the original terms (brown path). We then apply the ***Split*** rule to the original terms and to the terms produced by the ***Replace*** rule (green paths). The ***D0*** rule is applied to the original terms and to the terms produced by the ***Replace*** and ***Split*** rules (blue paths). Finally, we apply the ***D1*** rule to all the terms (pink paths).

For example, starting from the original term "Congenital adrenal gland hypoplasia", the rules ***Replace*** and ***Split*** do not produce any results. The modifier "Congenital" is removed by rule ***D0***, producing the demodified disorder term "adrenal gland hypoplasia". Finally, the modifier "hypoplasia" is removed by rule ***D1***, producing the term "adrenal gland", which corresponds to an anatomical structure.

Mapping demodified HPO terms to SNOMED CT concepts

Having removed the modifiers according to the transformation rules, we map all transformed terms to SNOMED CT through the UMLS using simple lexical mapping techniques [4]. More specifically, we attempt an exact match, followed by a normalized match using the functions provided by the UMLS Terminology Services API.

It is important to note that the first three rules (***Replace***, ***Split*** and ***D0***) are expected to produce more general terms for disorders, and yield mappings to disorder concepts in SNOMED CT (light blue links on Figure 1). In contrast, the ***D1*** rule produces terms for anatomical structures, physiologic processes or chemical substances. These terms are expected to map to entities of these types in SNOMED CT (orange link on Figure 1). For example, the term "factor XIII" (demodified from the HPO concept "Reduced factor XIII activity" [HPO:HP_0008357]) is mapped to the SNOMED CT concept "Factor XIII" [SNCTID: 319930009].

Of note, the original term and several demodified terms that derive from it may map to SNOMED CT. For example, "Congenital adrenal gland hypoplasia" (original HPO term), "adrenal gland hypoplasia" (produced by ***D0***) and "adrenal gland" (produced by ***D1***) all map to SNOMED CT. In this case, precedence is given to the mapping from the original term or to the mapping derived from the less aggressive transformation.

Creating post-coordinated expressions for HPO terms

In this preliminary work, we focus on expressions for D1 level modifiers, because D1 transformation rules tend to be more productive compared to D0 rules.

Using the templates suggested by existing SNOMED CT concepts, we generate post-coordinated expressions by inserting into the template the anatomical structure, physiologic process or chemical substance extracted by the transformation rule ***D1***. (An occasional D0 modifier may also have been removed from the term, but the final mapping happens at level D1.)

For example, the term "Macular" is extracted from the original HPO term "Macular hypoplasia" [HPO:HP_0001104] by the ***D1*** rule. It is mapped to the SNOMED CT concept "Macula lutea structure (body structure)" [SNCTID:362517001]. The modifier "hypoplasia" is associated with the template "Congenital hypoplasia of <ANATOMICAL STRUCTURE>", into which we insert the anatomical structure concept "Macula lutea structure (body structure)". The resulting logical definition for "Macular hypoplasia" is as follows.

'Disease (disorder)' *and ('Role group (attribute)' some (('Associated morphology (attribute)' some* ***'Hypoplasia (morphologic abnormality)'****) and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some* ***Macula lutea structure (body structure)'****)))*

Table 1 – List and count of modifiers organized in 4 levels: R (REPLACE), S (SPLIT), D0 and D1.

Description	n
Level R: Modifiers preventing terms from being mapped to disorders Modifiers for substitution patterns (one-to-one) *1st => first, 2nd => second, 3rd => third, 4th => fourth, 5th => fifth*	5 (present in 1,193 HPO terms)
Level S: Modifiers preventing terms from being mapped to disorders Modifiers for substitution patterns (one-to-two) Terms containing a "/" are split into two terms, for example: *Aplasia/Hypoplasia of kidney => Aplasia of kidney; Hypoplasia of kidney*	69 (present in 543 HPO terms)
Level D0: Modifiers preventing terms from being mapped to disorders *asymmetric, asymmetrical, bilateral, complete, congenital, cutaneous, generalized, lethal, marked, mildly, multiple, nearly, osseous, partial, patchy, severely, symmetric, symmetrical, unilateral, Alcohol-induced, Aminoglycoside-induced, Anesthetic-induced, Aspirin-induced, Cold-induced, Drug-induced, Effort-induced, Exercise-induced, Fava bean-induced, Heparin-induced, Radiation-induced, Stress-induced, infection-induced, Viral infection-induced, Warfarin-induced, ", recurrent", ", acute", ", chronic"*	37 (present in 1,129 HPO terms)
Level D1: Modifiers preventing terms from being mapped to anatomical structures/physiological processes/chemical substances *abnormal, abnormality, absence, absent, activity, agenesis, aplasia, aplastic, atresia[s], atrophy, bracket, broad, bullet shaped, bullet-shaped, chevron shaped, chevron-shaped, cone shaped, cone-shaped, contractures, curved, decreased, deficiency, degeneration, delayed, duplication, dystrophy, eeg, elevated, enlarged, fragmentation, hypoplasia, hypoplasic, hypoplastic, impaired, increased, irregular, ivory, loss, lytic defect[s], malrotation, number[s], osteolytic defect[s], prominent, pseudoepiphysis, reduced, rhomboid shaped, rhomboid-shaped, rudimentary, sclerosis, shortened, shortening, small, sparse, stippling, symphalangism, synostosis, thin, triangular, triangular shaped, triangular-shaped, wedge shaped, wedge-shaped, widened, widening*	65 (present in 8,864 HPO terms)

Results

Establishing a list of modifiers and transformation rules for HPO terms

As mentioned earlier, we identified four transformation rules (***Replace***, ***Split***, ***D0*** and ***D1***). As shown in Table 1, each rule is associated with a list of substitution patterns for replacement (***Replace***, ***Split***) or modifiers to be removed (***D0***, ***D1***). The number of patterns/modifiers per transformation rule ranges from 5 (***Replace***) to 69 (***Split***). The most frequent modifiers in HPO terms are those from the ***D1*** list (65 modifiers found in 8,864 HPO terms).

Identifying templates for phenotype concepts

As shown in Table 2, we identified 12 templates corresponding to modifiers at the ***D1*** level (with an occasional ***D0*** modifier), corresponding to five distinct logical definitions. Each template has the form:

{$modifier_1$, $modifier_2$, …, $modifier_n$} <ENTITY TYPE>

(e.g., {absence of}<ANATOMICAL STRUCTURE>)

Or : <ENTITY TYPE>{$modifier_1$, $modifier_2$, …, $modifier_n$}

(e.g., <ANATOMICAL STRUCTURE>{absent})

Each template is associated with a logical definition. Multiple templates can share the same logical definition. For example, the two templates listed above have the same logical definition. One of the authors (JTC) familiar with the SNOMED CT concept model inspected the templates for validity.

Applying transformation rules to HPO terms

The results of the successive application of transformation rules to HPO terms are depicted in Figure 1. The number of terms generated by each rule is listed next to the corresponding arrow. For example, 1,057 terms result from transformation ***D1*** applied to the 1,193 unique terms resulting from the ***Replace*** transformation (pink path ***R→D1***).

Mapping demodified HPO terms to SNOMED CT concepts

As we showed in our previous work [4], the mapping of the 16,413 original HPO terms to SNOMED CT through UMLS results in 3,081 HPO classes mapped to 4,215 SNOMED CT disorder concepts. Additionally, a total of 2,865 modified HPO terms (2,109 unique HPO classes) results from the three transformation processes (***Replace***, ***Split***, ***D0***). These disorder terms map to 515 SNOMED CT classes. The ***Replace*** transformation does not yield any new mappings. The ***Split*** transformation yields 55 mappings, of which 47 are new (i.e., not produced earlier). Transformation ***D1*** yields 2,857 mappings to anatomical structure, physiologic process or chemical substance concepts in SNOMED CT, 2,099 of which are new. The details of the mapping results can be found in Figure 1. Note that these mappings to pre-coordinated concepts in SNOMED CT only contribute to the post-coordination strategy developed in this paper.

Creating post-coordinated expressions for HPO terms

In this proof-of-concept investigation, we identified 12 templates involving 10 ***D1*** modifiers (and 3 additional ***D0*** modifiers removed when necessary). The instantiation of the logical definitions associated with these templates covers 1,617 HPO classes. Compared to the 3,081 mappings to pre-coordinated SNOMED CT concepts, this approach increases the coverage of HPO classes in SNOMED CT by roughly 50% through post-coordination.

Discussion

Our approach is very productive, yielding substantial gain in terms of number of mappings. More specifically, we developed post-coordinated expressions for 1,617 HPO classes that were not previously mapped to SNOMED CT classes, increasing the total number HPO classes mapped to SNOMED CT by 50%.

A high level of implicit knowledge was observed in HPO. For example, the notion of congenitality is usually assumed, rather than stated in HPO terms. While understandable in the context of clinical genetic phenotypes, it hinders the mapping to other terminologies. In contrast, SNOMED CT uses the qualifier value "congenital" to explicitly denote congenitality. Moreover, the finer-grained definitions supported by post-coordination make it possible to represent laterality (unilateral, bilateral, left, right), which is generally not represented in pre-coordinated terminologies.

Some of the templates we created do not fit the SNOMED CT concept model (e.g., {EEG with} <PHYSIOLOGICAL PROCESS> for electroencephalogram waveforms description) or lack specific pre-coordinated classes (e.g., for specific enzymes {decreased activity of} <CHEMICAL SUBSTANCE>). Additional work is needed to explore how these templates could be represented in SNOMED CT.

In future work, we want to identify templates for ***D0*** modifiers. We also want to assess the logical validity of the post-coordinated expressions we created by classifying them with the description logics classifier used by SNOMED CT. Finally, we would like to test the applicability of this method for mapping other specialized terminologies to SNOMED CT.

Conclusion

In this preliminary study, we explored the automatic mapping of HPO terms to SNOMED CT through post-coordination. Through this novel approach, we were able to increase the current number of mappings by 50%.

References

[1] Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc JAMIA. 2013 Jun;20(e1):e147–54.

[2] Hennekam RCM, Biesecker LG. Next-generation sequencing demands next-generation phenotyping. Hum Mutat. 2012 May;33(5):884–6.

[3] Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, et al. The PhenX Toolkit: get the most from your measures. Am J Epidemiol. 2011 Aug 1;174(3):253–60.

[4] Winnenburg R, Bodenreider O. Coverage of Phenotypes in Standard Terminologies. Proceedings of the Joint Bio-Ontologies and BioLINK ISMB'2014 SIG session "Phenotype Day." Boston, USA; 2014. p. 41–4.

[5] Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014 Jan;42(Database issue):D966–74.

[6] Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform. 2006;121:279–90.

[7] Rector A, Iannone L. Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. J Biomed Inform. 2012 Apr;45(2):199–209.

[8] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267–70.

[9] Groza T, Hunter J, Zankl A. Decomposing Phenotype Descriptions for the Human Skeletal Phenome. Biomed Inform Insights. 2013 Feb 4;6:1–14.

[10] Oellrich A, Grabmüller C, Rebholz-Schuhmann D. Automatically transforming pre- to post-composed phenotypes: EQ-lising HPO and MP. J Biomed Semant. 2013 Oct 16;4:29.

[11] Sollie A, Sijmons RH, Lindhout D, van der Ploeg AT, Rubio Gozalbo ME, Smit GPA, et al. A new coding system for metabolic disorders demonstrates gaps in the international disease classifications ICD-10 and SNOMED-CT, which can be barriers to genotype-phenotype data sharing. Hum Mutat. 2013 Jul;34(7):967–73.

[12] Beck T, Free RC, Thorisson GA, Brookes AJ. Semantically enabling a genome-wide association study database. J Biomed Semant. 2012;3(1):9.

Acknowledgments

This work was supported in part by the Intramural Research Program of the NIH, National Library of Medicine, the French Gynecology and Obstetrics Association (Collège National des Gynécologues et Obstétriciens Français), and the Philippe Foundation.

Address for correspondence

Corresponding author: olivier@nlm.nih.gov

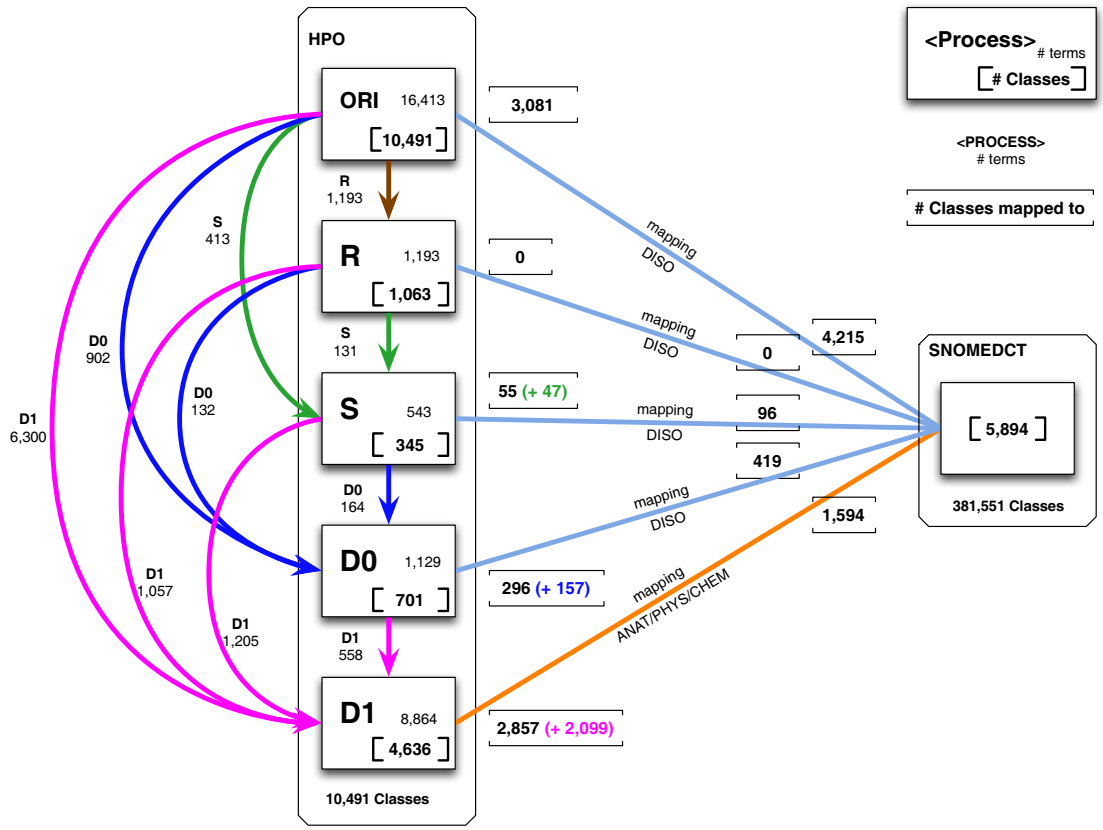


Figure 1 – Levels of transformations for the 5,384 mappings between HPO and SNOMEDCT Classes (ORI : original HPO terms; R=Replace transformation; S=Split transformation; D0/D1= Demodification transformations)

Table 2 – Templates and logical definitions for post-coordinated expressions (D1 level).

Template	Logical definition	# Classes (terms)
{abnormal, abnormality of, abnormality of the, abnormality involving the }<ANATOMICAL STRUCTURE> <ANATOMICAL STRUCTURE>{abnormal}	'Disease (disorder)' and ('Role group (attribute)' some (('Associated morphology (attribute)' some 'Developmental anomaly (morphologic abnormality)') and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some <ANATOMICAL STRUCTURE>)))	618 (714)
{aplastic, aplasia of, aplasia of the, aplasia involving the, absence of}<ANATOMICAL STRUCTURE> ANATOMICAL STRUCTURE>{agenesis, aplasia, absent} {congenital absence of, congenital aplasia of }<ANATOMICAL STRUCTURE>	'Disease (disorder)' and ('Role group (attribute)' some (('Associated morphology (attribute)' some 'Congenital absence (morphologic abnormality)') and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some <ANATOMICAL STRUCTURE>)))	415 (977)
{hypoplastic, hypoplasia of, hypoplasia of the, hypoplasia involving, hypoplasia involving the, hypoplasia affecting the}<ANATOMICAL STRUCTURE> <ANATOMICAL STRUCTURE>{hypoplasia} {congenital hypoplasia of}<ANATOMICAL STRUCTURE>	'Disease (disorder)' and ('Role group (attribute)' some (('Associated morphology (attribute)' some 'Hypoplasia (morphologic abnormality)') and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some <ANATOMICAL STRUCTURE>)))	409 (853)
{duplication of, duplication of the, duplication involving}<ANATOMICAL STRUCTURE> <ANATOMICAL STRUCTURE>{duplication} {complete duplication of, complete duplication of the}<ANATOMICAL STRUCTURE>	'Disease (disorder)' and ('Role group (attribute)' some (('Associated morphology (attribute)' some 'Double structure (morphologic abnormality)') and ('Occurrence (attribute)' some 'Congenital (qualifier value)') and ('Finding site (attribute)' some <ANATOMICAL STRUCTURE>)))	140 (232)
{bilateral, D1} <ANATOMICAL STRUCTURE> examples : {bilateral aplasia}<ANATOMICAL STRUCTURE> {bilateral}<ANATOMICAL STRUCTURE>{aplasia} {bilateral absence of}<ANATOMICAL STRUCTURE> …	'Disease (disorder)' and ('Role group (attribute)' some ***… D1 logical definition here …*** and ('Finding site (attribute)' some 'left <ANATOMICAL STRUCTURE>')) and ('Role group (attribute)' some ***… D1 logical definition here …*** and ('Finding site (attribute)' some 'right <ANATOMICAL STRUCTURE>'))	35 (63)
	TOTAL =	1,617 (2,839)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-800

A Model-Driven Approach to Customize the Vocabulary of Communication Boards: Towards More Humanization of Health Care

Natália M. Franco[a,b], Gabriel F. Medeiros[a], Edson A. Silva[a], Angela S. Murta[a], Aydano P. Machado[b], Robson N. Fidalgo[a]

[a] *Informatics Center, Federal University of Pernambuco, Recife, Brazil*
[b] *Computing Institute, Federal University of Alagoas, Maceió, Brazil*

Abstract

Objective: This work presents a Modeling Language and its technological infrastructure to customize the vocabulary of Communication Boards (CB), which are important tools to provide more humanization of health care.

Method: Using a technological infrastructure based on Model-Driven Development (MDD) approach, our Modeling Language (ML) creates an abstraction layer between users (e.g., health professionals such as an audiologist or speech therapist) and application code. Moreover, the use of a metamodel enables a syntactic corrector for preventing creation of wrong models.

Results: Our ML and metamodel enable more autonomy for health professionals in creating customized CB because it abstracts complexities and permits them to deal only with the domain concepts (e.g., vocabulary and patient needs). Additionally, our infrastructure provides a configuration file that can be used to share and reuse models. This way, the vocabulary modeling effort will decrease over time since people share vocabulary models.

Conclusion: Our study provides an infrastructure that aims to abstract the complexity of CB vocabulary customization, giving more autonomy to health professionals when they need customizing, sharing and reusing vocabularies for CB.

Keywords:

Model-Driven Development; Modeling Language; Metamodel, eHealth, Assistive Technology; Communication Board; Speech Impairment.

Introduction

People have speech problems for many reasons (e.g., autism, neck injury, ASL, cerebral palsy, stroke, dysarthria, and tracheostomy) and they communicate their needs, desires and complaints in an alternate way. The absence of this communication causes lack of thinking, frustration, anger, and social exclusion of these individuals [1,2]. According to Palerm & Ruiz [3], until six years old, children have their cognitive development directly related to language development. The lack of communication with other people, in this stage of life, can irreversibly damage their intellectual capacity.

In this context, Augmentative and Alternative Communication (AAC) [4] is a field of Assistive Technology [5] dedicated to expanding the communication skills of these people. According to the American Speech-Language-Hearing Association (ASHA) [4], AAC is the area of clinical practice that attempts to compensate for difficulties or disabilities demonstrated (either temporarily or permanently) by individuals with severe communicative disorders. AAC resources recognize and value all communication attempts of an individual – whether through gestures, facial expression, eye gaze, writing or drawings – and try to complement them.

In the AAC context, Communication Board (CB) (Figure 1) is one of the best ways to address basic needs of expression, because the CB is a set of symbols and captions that represent objects, actions, feelings, etc. Therefore, the communication is established using a selection of symbols and captions that express what the user wants to say. Generally, they are arranged into groups and colors in order to facilitate symbol location tasks. The most used pattern is the Fitzgerald Key [6] notation, where the vocabulary is divided into six colored groups as follows: 1) Nouns in orange; 2) Pronouns in yellow; 3) Verbs in green; 4) Adjectives in blue; 5) Social Expressions in pink; and 6) Miscellany in white.

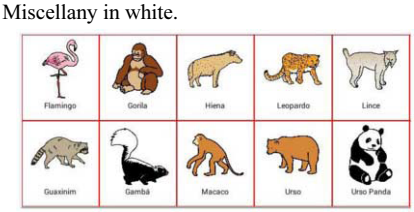


Figure 1 – Communication Board example

The universe of eHealth technologies to facilitate communication is increasing and can now help several speech-impaired people to communicate and socialize with others. According to the theory of communication [7], effective communication can only occur between people (i.e., sender and receiver) who understand the code (i.e., vocabulary) and can code (sender) and decode (receiver) the message. This way, the vocabulary is a key concept in communication, and it is unique for each person.

Despite all eHealth potential benefits of CB, most of the solutions do not allow vocabulary customization for a given patient. Most of them come with a huge set of default symbols which is more than necessary for effective communication, and causes cognitive overload in the user (e.g., many options to choose from in the search area). The CB eHealth solutions that allow vocabulary customization do so by filling out fields in a Graphical User Interface (GUI) [4, 5]. This approach has some problems, for example, since each element is separately built, it is not possible to create a cohesive relationship among them. Thus, in this scenario, a patient using a CB can make all possible combinations with CB vocabulary elements, some of them causing communication cohesion problems. For instance, it is possible for communicating something like *"I want to wear an apple"*. This information is not consistent. The verb *"wear"* needs an outfit as a complement, and the noun *"apple"*, a verb related to food. Moreover, the vocabu-

lary customization made from the fulfillment of fields in a GUI makes the job of users harder (e.g., audiologist or speech therapist) because there is no global overview of the created content. This way, the user can mistakenly insert the same word more than once, for example.

In order to propose a better eHealth solution for CB vocabulary customization and design a viable technology for this task, it is worth thinking about some questions: Is there a way to express our vocabulary organization mental model and discuss it with another person? What methods are there to abstract the complexity of this process? There is a domain-specific solution? How can eHealth improve this process? Is there a standard for CB vocabulary customization? How can developers adopt reusability of the vocabulary? Can productivity be enhanced? What about interoperability?

After thinking about those questions and aiming to overcome the previous shortcomings (i.e., cognitive overload, no cohesion relationship and no global vision), we have developed a new approach, based on Model-Driven Development (MDD) [10], visual Modeling Language (ML) [11] and metamodel [12] in order to customize CB vocabulary. Many successful solutions for eHealth [9, 10] use MDD. This choice can be explained by the main characteristic of MDD, the creation of an abstraction layer to deal with complexity. Moreover, we can process visual information efficiently, about a quarter of our brain is devoted to vision, more than all our other senses combined [15]. Information represented visually is also more likely to be remembered due to the picture superiority effect [16]. As a consequence, our proposal tends to be more intuitive and easy, giving autonomy to health professionals in performing these tasks.

In this context, the main objective of this paper is to present how Software Engineering, more specifically Model-Driven Development (MDD) based on Modeling Language (ML) and metamodel, can be used to improve communication and quality of software programming for Augmentative and Alternative Communication (AAC). Additionally, we discuss how these technologies can improve health humanization.

Materials and Methods

Model-Driven Development

Model-Driven Development (MDD) is defined as a software development paradigm which uses models as the key artifacts in all development phases. For instance, a Modeling Language (ML) is a visual notation and it is defined by an abstract syntax (i.e., grammar), a concrete syntax (i.e., representation or notational elements), mappings between abstract and concrete syntax, and a description of the semantics [12].

In MDD, models are more than documentation items, they are executable objects [12]. In other words, once the models are created, target code can be generated and then compiled or interpreted for execution. Thus, MDD and ML hide implementation details, which speeds up the software development process by reducing its complexity [12]. Moreover, visual notations tend to provide higher-level abstraction than textual ones, because it is domain-specific. In this way, ML tends to be easier than pure coding, especially because the ML visual notation aims to deal only with the domain concepts and abstract textual representation, hiding implementation details.

MLs are supported by Computer-Aided Software Engineering (CASE) tools that can detect errors earlier, or even prevent them from happening, guide towards preferable design patterns, check completeness by informing about missing elements, reduce modeling work by applying conventions and default values, support full code generation, and keep specifications consistent [12].

The specification of a ML in MDD is based on the metamodel concept [12], which provides an abstract syntax to distinguish valid and invalid models. That is, a metamodel (or abstract syntax) of a ML describes concepts, relationships between them, and structuring rules that constrain the combination of these concepts according to the domain rules of a ML. Thus, a metamodel presents a precise specification of modeling concepts and well-formed rules needed for creating syntactically correct models. Consequently, a metamodel is as useful for a ML as grammar is for a programming language.

In this context, it is important to highlight that a metamodel for ML offers the basic definitions (i.e., metadata) to develop CASE tools, which can be done using free (e.g., Eclipse Modeling Project [17]) or proprietary (e.g., Microsoft Modeling SDK for Visual Studio [18]) technologies.

Our Proposal

Figure 2 illustrates an overview of our eHealth proposal for CB vocabulary customization using MDD. The proposal consists of two components: 1) A CASE Tool for modeling the CB vocabulary; and 2) The CB Generator for generating an application (*app*) for a touch screen device.

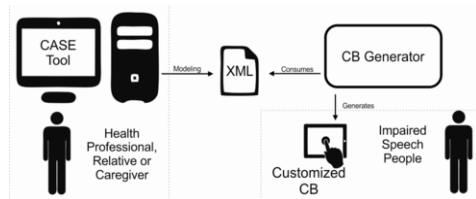


Figure 2 – Our proposal overview

This paper focuses on the use of the MDD paradigm for building an infrastructure that allows health professionals (e.g., audiologist or speech therapist) to easily customize CB vocabulary for each patient.

In order to build this infrastructure, we need to define our application domain, that is the CB vocabulary universe. Following this, we need to define the ML (i.e., concrete syntax and abstract syntax) for this domain. Our ML, named Communication Board Modeling Language (CBML), is described in the following.

The concrete syntax (i.e., notational elements) was defined based on the original Fitzgerald Key notation [6]. Such notation was chosen because it is a widely used pattern for professionals (e.g., audiologist or speech therapist) in CB eHealth solutions. The abstract syntax (i.e., metamodel) is elaborated after the concrete syntax and defines the domain rules, that means, how to connect the notational elements to build correct models. Our ML was defined to allow vocabulary categorization and to limit constructions with cohesion problems. The CBML (i.e., notational elements and metamodel) will be further discussed in the Results section.

With these previous artifacts (i.e., notational elements and metamodel) the software (i.e., CASE Tool and CB Generator) can be implemented. The CASE Tool is based on a desktop computer and its users are health professionals, familiars or caregivers. In turn, the CB Generator is based on a mobile device with a touch screen (e.g., tablets and smartphones) and its users are people with oral communication impairments. These two modules communicate through a file in an Extensible Markup Language (XML) [19] format that is generated from the model designed in the CASE tool. This XML file contains all of the vocabulary modeled in the CASE tool and

serves as an input to the CB Generator to customize the CB. The CASE Tool was discussed in another paper [20] and a brief discussion of CB Generator is shown in the Results section.

It is important to mention that CB customization (e.g., contrast color, size and number of icons per screen) and its usability requirements (e.g., sonorous feedback and scanning mode) are the responsibility of the CB Generator. The ML only defines the CB vocabulary.

Results

CBML

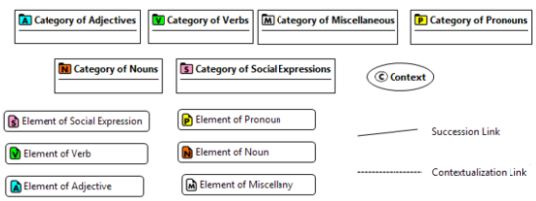


Figure 3 – CBML notational elements

Figure 3 shows the CBML elements, based on the original Fitzgerald Key notation [6]. In this work, the white color is also used to represent a *Context* concept. The notational elements of CBML are:

1. *Categories*, represented by a white rectangle with a folder pictogram with the initial letter of group name and color, as specified by the Fitzgerald Key pattern;
2. *Elements*, represented by a white rectangle with a file pictogram with the initial letter of group name and color, as specified by the Fitzgerald Key pattern;
3. *Context*, represented by a white ellipse with a circle pictogram with the letter *"C"*;
4. *Succession Link*, represented by a continuous line;
5. *Contextualization Link*, represented by a dashed line.

Regarding pictograms, a *Category* is represented by a folder icon that can contain other folders (subcategories) and several *Elements*, which are represented by file icon.

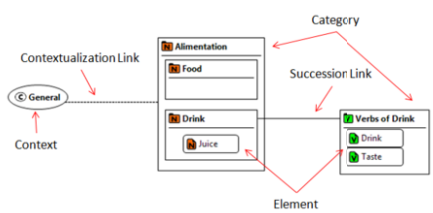


Figure 4 – CBML model example (node-link diagram)

Figure 4 demonstrates how CBML notational elements can be organized. In this example, the concept *"Alimentation"* was categorized into *"Food"* and *"Drink"*, and the subcategory *"Drink"* has the *Element "Juice"*. The *Category "Verbs of Drink"* has the *Elements "Drink"* and *"Taste"*, and it is connected to the *"Drink" Category* through a *Succession Link*. The *"General" Context* is connected to *"Alimentation"* through a *Contextualization Link*. Due to the *Succession Link*, a semantic relationship is created between words. This way, the user cannot build sentences with cohesion problems (e.g., *"drink a food"*), because the verb *"Drink"* is connected only to the *"Drink" Category*.

Metamodel CBMM

The Communication Board MetaModel (CBMM) is designed considering that the CB vocabulary model is essentially a node-link diagram [21], where a CB vocabulary diagram is composed of nodes representing *Categories*, *Elements* and *Contexts*, and links representing *Succession Links* and *Contextualization Links*. In Figure 4 there are some nodes and links of the CB vocabulary model. A conceptual view of our metamodel is presented in Figure 5.

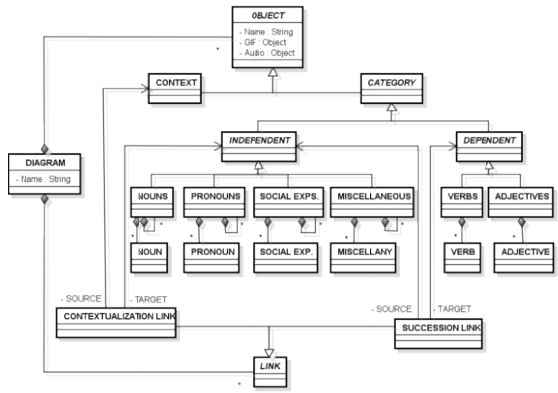


Figure 5 – The CBMM metamodel

In CBMM there are three main meta-concepts: *Diagram*, *Object* and *Link*. *Diagram* is the root meta-concept that corresponds to the drawing area of a CB diagram. For this reason, *Diagram* can have many instances of *Object* and many instances of *Link*, which cannot exist without a *Diagram*. An *Object* corresponds to an icon in the CB GUI and it is characterized by a caption, an image, and an audio, which are represented by *Name*, *GIF* and *Audio* attributes. Alongside these main meta-concepts, our metamodel has two specialized meta-concepts of *Object*, *Category* and *Context*, which are described as follows.

The *Category* meta-concept is responsible to group concepts according to the Fitzgerald Key notation, and it is specialized in two meta-concepts, *Dependent* and *Independent*, which are presented as follows.

On one hand, the *Dependent* meta-concept is responsible for organizing those concepts of the Fitzgerald Key that can generate cohesion problems (e.g., *"eat a juice"* and *"soft milk"*). It is specialized into two meta-concepts, *Verbs* and *Adjectives*. On the other hand, the *Independent* meta-concept is responsible for organizing the other concepts and it is specialized in four meta-concepts: *Nouns*, *Pronouns*, *Social Expressions* and *Miscellaneous*. Note that each meta-concept of the Fitzgerald Key notation is referenced in plural form (e.g., *Nouns*, *Pronouns*) and they are associated with the notational element of *Category* (Figure 3). These meta-concepts, in plural form, can be recursively composed to create a subcategory in a *Category* (except *Verbs* and *Adjectives*). Additionally, each meta-concept in its plural form is composed of many instances of another meta-concept, in singular form, that is associated with the notational element of an *Element* (Figure 3).

The *Context* meta-concept, specialized from the *Object* meta-concept, is related to the environment concept (e.g., *"bed"* in a *"domestic context"* and *"stretcher"* in a *"hospital context"*).

In sequence, the *Link* meta-concept is specialized into *Contextualization Link* and *Succession Link*. In the first case, the *Contextualization Link* is responsible to connect a *Category* of concepts to a *Context*. In this way, the *Contextualization Link*

has a *source* relationship with a *Context* and a *target* relationship with an *Independent Category*. In the second case, the *Succession Link* is responsible to avoid communication cohesion problems. In this way, the *Succession Link* has a *source* relationship with an *Independent Category* and a *target* relationship with a *Dependent Category*.

CASE Tool and Front-end Application

In order to demonstrate the feasibility, expressiveness, and usefulness of our ML and metamodel, we have developed the Communication Board CASE (CBCASE), that was discussed in another paper [20]. The front-end application is in development and we already have a low-fidelity prototype (Figure 6).

Figure 6 – Front-end interface (low-fidelity prototype)

The main idea is the use of a grid format in landscape mode with a toolbar at the bottom. The main objective of the grid format is to facilitate use by people with motor impairment through the use of Assistive Technologies as scanning mode and binary switches [22]. This allows people to use our CB without the need to touch it. For instance, the end user can easily access the main functions and options such as navigation, fast answers, and favorite expressions right from the toolbar.

Discussion

Communication, in its various forms, is a basic human need for people to socialize, express ideas, feelings and desires. The absence of this necessity causes frustration, anger and social exclusion of these individuals. A study with patients who received mechanical ventilation during their period in an intensive care unit noted that the level of frustration in trying to communicate decreased with the insertion of a CB [1]. We know emotions influence the physical and psychological health of the patient. In this case, it was observed they were calmer when they could effectively communicate their complaints, sensations and feelings.

In this context, our study aims to address three main objectives: 1) achieve health humanization by enabling dialogs between health professionals and patients; 2) provide an environment to allow health professionals (e.g., audiologist or speech therapist) to easily customize, share, discuss and reuse CB vocabulary; and 3) offer a CB with customized vocabulary for each patient, giving voice, autonomy and quality of life to them.

Vocabulary is the communication core and depends on several factors (e.g., age, culture, location, and context) in such a way that it is customized for each person. The current eHealth solutions for CB, if they allow vocabulary customization, do this by filling out the fields in a GUI. This approach has some problems, for example, no global vision of created content and cohesion problems during use. This customization problem could be solved by using an ontology editor (e.g., Protegé). However it is a generic editor (i.e., it is not domain-specific) that is too complex to be used, even for computer professionals, thus making the modeling work a difficult task.

With the MDD paradigm it is possible to create an abstraction layer between the user who will create the CB vocabulary (e.g., health professional as audiologist or speech therapist) and the implementation code. Using ML the user only needs to know about the specific context (e.g., vocabulary and patient needs). This way, the user, who deeply knows about the patient environment and needs, can do it on their own without depending on programmers.

The CBML allows the CB vocabulary modeling in a diagrammatic way, meaning it can be done, for example, with a pencil on a piece of paper. This feature enables discussion among users to improve the vocabulary and, when it is done, the user can model it in the CASE tool for generating the application code. On the other hand, our architecture provides an infrastructure to encourage people to share vocabularies through XML files. This way, the effort for modeling a vocabulary will decrease with time as XML files are shared and reused. Moreover, these XML files can be used as a standard for CB vocabulary customization, enabling interoperability among different platforms.

Additionally, our solution (i.e., CBML, CBMM, CBCASE and the front-end application) prevents wrong models and cohesion problems because it is based on a metamodel. That is, a metamodel (or abstract syntax) of a ML describes the concepts, the relationships between them, and the structuring rules that constrain the combination of these concepts according to the domain rules of a ML. In other words, a metamodel is useful to provide a precise definition of all modeling concepts and well-formed rules needed for creating syntactically correct models [12]. In this context, we enumerate the main well-formed rules that are intrinsically assured by our metamodel (Figure 5). That is, these well-formed rules are directly derived from our metamodel and, for this reason, can be checked against it. These rules are:

1. An *Object* (e.g., *Category* and *Context*) cannot be connected to another *Object* without a *Link*, and vice-versa. That is, neither an *Object* can be directly connected to other *Objects* nor a *Link* can be directly connected to other *Links*;
2. A *Link* cannot be created without an *Object* as a source and another as a target. That is, a *Link* cannot be depicted isolated in the drawing area;
3. A *Subcategory* can only be built within another *Category* of the same type (e.g., nouns, pronouns, social expressions and miscellaneous);
4. An *Element* can only be instantiated within a *Category* of the same type (e.g., nouns, pronouns, social expression, miscellaneous, verbs and adjectives);
5. A *Succession Link* can only exist between *Dependent Categories* (e.g., verbs and adjectives) and *Independent Categories* (e.g., nouns, pronouns, social expressions and miscellaneous);
6. A *Contextualization Link* can only exist between *Context* and *Independent Categories* (e.g., nouns, pronouns, social expressions and miscellaneous).

Conclusion

Communication is an indispensable factor to address health humanization. Both aphasic patients and health professionals need to be heard, establishing a network of dialogs to lower distances, connecting people and increasing patient quality of

life. Current solutions, in allowing vocabulary customization, do it by filling out fields in a GUI. This approach causes problems such as cognitive overload, no cohesion relationship in message, and no global vision of the modeled content. Consequentially, it makes the job of health professional who will model the vocabulary harder, therefore generating a CB with low usability for the user.

In this paper, we presented the use of the MDD paradigm for eHealth, especially for CB, which aims to help people unable to use speech to communicate. In this context, MDD provides an abstraction layer to enable health professionals to easily customize a vocabulary for each patient.

We presented CBML, the ML to diagram and discussed the vocabulary, and CBMM, the metamodel that describes the concepts and how to combine them according to the domain rules. CBML and CBMM, together, provide a new paradigm for the CB vocabulary customization. Our proposal is original and provides a way to organize a vocabulary as our mental model. Another point to highlight in our solution is the possibility to create a new standard for CB vocabulary customization. It can be made through our configuration file, enabling interoperability among platforms. Furthermore, using the MDD approach, it is possible to work directly with domain concepts, thus facilitating learning and use.

To demonstrate the feasibility, expressiveness and usefulness of CBML and CBMM, we have developed the CBCASE that was discussed in another paper [20]. The main contribution of this paper is the MDD approach, using ML and metamodel, that can be used as a standard for CB vocabulary customization.

In conclusion, our work contributes to the advancement of the state of the art of MDD, ML and metamodel for CB; it provides a novel perspective of scientific research and industrial application in the field of vocabulary customization for a given patient. Furthermore, our metamodel is the first one that aims to customize a CB vocabulary. We highlight three main features of our ML that are guaranteed by our metamodel: 1) *Category*, to provide vocabulary modularization; 2) *Succession Link*, to avoid cohesion problems (e.g., *"wear an apple"*); and 3) *Contextualization Link*, to filter the vocabulary according to the situation (e.g., home, hospital, school).

Acknowledgements

This research was supported by the National Center for Scientific and Technological Development (CNPq) and the Coordination of Improvement of Higher Education Personnel (CAPES).

References

[1] Patak, L.; Gawlinski, A.; Fung, N. I.; Doering, L.; Berg, J. and Henneman, E. A. *Communication boards in critical care: patients' views*. Appl. Nurs. Res., vol. 19, no. 4, pp. 182–190, 2006.

[2] Bahia, M. M. and Chun, R. Y. S. *Quality of life in aphasia: differences between fluent and non-fluent aphasic Augmentative and Alternative Communication users*. Audiol. - Commun. Res., pp. 1–8, 2014.

[3] Palerm, C. C. R. and Ruiz, P. A. L. *Augmentative / Alternative System for Public Communication for Children with Cerebral Palsy* in 14th Annual International Conference of Engineering in Medicine and Biology Society of the IEEE, 1992, pp. 1525–1526.

[4] ASHA. *American Speech-Language-Hearing Association*. http://www.asha.org/. (last accessed: 30 Oct 2014)

[5] Secretaria Especial dos Direitos Humanos. Comitê de Ajudas Técnicas. *Tecnologia Assistiva*. Brasília: CORDE, 2009, p. 138.

[6] Fitzgerald, E. *Straight Language for the Deaf. A System of Instruction for Deaf Children*. Alexander Graham Bell Association for the Deaf, 1949, p. 104.

[7] Moody, D. L. *The 'Physics' of Notations: Towards a Scientific Basis for Constructing Visual Notations in Software Engineering.* vol. 35, no. 5, pp. 756–778, 2009.

[8] Que-Fala. *Que Fala!*. http://metodos-si.com.br/. (last accessed 30 Oct 2014)

[9] Mayer-Johnson. *BoardMaker Software*. http://www.mayer-johnson.com/boardmaker-software/. (last accessed 30 Oct 2014)

[10] Brambilla, M.; Cabot, J. and M. Wimmer. *Model-Driven Software Engineering in Practice*. Morgan & Claypool Publishers, 2012.

[11] Hyun Cho, E. and Gray, J. and Syriani, *Creating visual Domain-Specific Modeling Languages from end-user demonstration.* In Modeling in Software Engineering (MISE), 2012 ICSE Workshop on, 2012, pp. 2–28.

[12] Kelly, K. and Tolvanen, J. P. *Domain-Specific Modeling: Enabling Full Code Generation*. Wiley, 2008.

[13] Raghupathi, W. and Umar, A. *Exploring a model-driven architecture (MDA) approach to health care information systems development.* Int. J. Med. Inform., vol. 77, no. 5, pp. 305–14, May 2008.

[14] Khambati, A.; Grundy, J.; J. Warren and Hosking, J. *Model-Driven Development of Mobile Personal Health Care Applications.* 2008 23rd IEEE/ACM Int. Conf. Autom. Softw. Eng., no. ASE, pp. 467–470, Sep. 2008.

[15] Kosslyn, S. M. *Graphics and Human Information Processing*. J. Am. Stat. Assoc., vol. 80, pp. 499–512, 1985

[16] Goolkasian, P. *Pictures, Words and Sounds: From Which Format Are We Best Able to Reason?.* J. Gen. Psychol., pp. 439–459, 2000.

[17] The Eclipse Foundation. *Eclipse Modeling Framework Project.* http://www.eclipse.org/modeling/emf/?project=emf. (last accessed: 30 Oct 2014).

[18] Microsoft. *Visual Studio.* http://www.visualstudio.com/. (last accessed: 30 Oct 2014).

[19] W3C. *Extensible Markup Language (XML).* http://www.w3.org/XML/. (last accessed: 30 Oct 2014).

[20] Franco, N. M.; Fidalgo, R. N.; Silva, E. A.; Cavalcante, T. F. and Brito, P. H. S. *Modeling Language and CASE Tool for Communication Board Customization* in Proceedings of the IEEE HealthCom 2014, 2014, pp. 277–282.

[21] Irani, P.; Tingley, M. and Ware, C. *Using Perceptual Syntax to Enhance Semantic Content in Diagrams.* In IEEE Computer Graphics and Applications, 2001, pp. 76–85.

[22] Stephanidis, C. and Savidis, A. *Universal Access in the Information Society: Methods, Tools, and Interaction Technologies*. Universal Access Informaton Society. pp. 40–55, 200.

Address for correspondence

Center of Informatics – Federal University of Pernambuco. Av. Professor Luís Freire, s/n – CDU – 50740-540, Recife, PE, Brazil. Email: nmf@cin.ufpe.br | rdnf@cin.ufpe.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-805

Pain Documentation: Validation of a Reference Model

Emily Gesner [a] Sarah A Collins[a,b,c], Roberto Rocha[a,b,c]

[a] *Clinical Informatics, Partners eCare, Partners Healthcare Systems, Boston, MA*
[b] *Department of General Internal Medicine, Brigham and Women's Hospital, Boston, MA*
[c] *Harvard Medical School, Boston, MA*

Abstract

Over the last decade, interoperability of the Electronic Health Record (EHR) is becoming more of a reality. However, inconsistencies in documentation such as pain are considered a barrier to obtaining this goal. In order to be able to remedy this issue, it is necessary to validate reference models that have been created based upon requirements defined by Health Level 7 (HL7), Logical Names and Codes (LOINC) and the Intermountain Clinical Element Model using external published sources and guidelines. Using pain as an example of complex and inconsistent documentation, it was found that the reference model based upon these standards is valid because the data elements identified are broad and can meet the needs of each sub-domain within the primary domain of pain.

Keywords:

Pain documentation, clinical reference model.

Introduction

Converting a hospital system that has been using various electronic health records (EHR) to a single electronic medical record (EMR) can be a significant challenge. Variation in clinical workflows may serve as rationales to request modifications to the user interface (UI) of an EHR, including data terminology and definitions, for specific settings or specialty care areas. While a goal of these modifications may be to increase productivity by providing tailored screen interactions for end-users, the integrity and consistency of data capture across an organization suffers. The creation of a reference data model and validation based upon external published references is critical to steer consistent data capture in any organization. "Data standards, including terminologies and common data elements (CDE), is a critical first step towards achieving automated data integration."[1]

This study is focused on Partners HealthCare (Partners), a large integrated health care delivery system in the Northeast region of the United States that consists of 12 hospitals that have been affiliated or acquired over the last 20 years. Each hospital has different inpatient and outpatient EHR's, which are either internally developed or vendor provided. The Partners eCare (PeC) Program is an enterprise-wide effort to implement a single vendor-based EHR[9]. The opportunity to convert all the hospitals to one EHR platform has given way to the need to standardize documentation throughout all the specialties. As the United States and other countries increase their infrastructures and capabilities to share electronic healthcare data, consistent data definitions based on clinical reference models will be critical.[2] If clinical data is inconsistently defined and captured, an important goal is to iteratively identify and resolve these data inconsistencies, starting while the new EHR platform is being configured, but continuing throughout the system deployment phases. Ideally, this iterative process should start with high priority clinical documentation topics.

Complexity of Pain Documentation

Pain documentation is an example of an important and complex clinical documentation topic. The complexity of pain documentation is derived from its broad application to a variety of clinical settings, specialty specific requirements, and multiple pain subtopics, such as cancer-related pain or acute post-procedural pain related to a joint replacement. The management of pain requires complex care coordination, clinical decision-making, and trending and evolution of the patient's response to treatment[3]. As complexity of care delivery increased, so does the complexity of documentation to reflect that care. Even educated, experienced medical professionals may rely on past personal experiences, outdated teachings, and be more resistant to incorporating new practice pain management guidelines which can directly affect the degree of documentation completed.[4] Therefore, consistency in application of evidenced-based pain management practice is the primary means to combat under treatment of pain.[5] Standardizing the complexities of structured data capture for pain documentation is hypothesized to increase continuity of care, improve the integrity and efficiency of secondary data use, and ultimately improve patient care. The first step is to review an existing reference model and validate it using external published sources.

At Partners, we have been identifying high priority clinical documentation topics, such as pain, to target for development of reference models. This work is performed as part of the PeC Program and includes ongoing collaborations between clinical experts, EHR analysts, and informaticians. Reference models are defined using a combination of sources, including Health Level 7 (HL7)[6], Intermountain's Clinical Element Model[7] and Logical Observation Identifiers Names and Codes (LOINC)[8]. However, implementing this reference model based on these requirements alone is not valid without the support of published external documentation. It is necessary to validate the model within each type of clinical setting and specialty specific requirements.

We describe in this paper the process we used to validate the reference model for pain, taking into account published evidence from a various sources.

Table 1- Acute Pain Reference Model

Reference Model Elements[1]	Elements derived from published sources[1]	Validation Outcome[2]
Pain Onset	Pain History	*V*
Pain Onset (Hours ago)	Pain History	*V*
Speed of Pain Onset	Pain History	*V*
Pain Primary Location	Location	*V*
Pain Quality (Character)	Quality of Pain	*V*
Pain Periodicity		
Pain Temporal Pattern		
Pain Alleviating Factors	Alleviating Factors	*V*
Pain Aggravating Factors	Aggravating Factors	*V*
Pain Duration (Hours, Minutes)	Pain Duration	
Associated Signs and Symptoms	Physiological and behavioral Responses to Pain	*V*
Patient Severity Score	Intensity of Pain	*V*
Pain Course	Pain Treatment	*V*
Relative Temporal Context	*M*	*MN*
Patient Stated Goal	*M*	*MN*

1 M= Missing
2 V= Validated, MN= Missing and Not Required, MR= Missing and Required

Methods

Literature Review of Pain Documentation Guidelines

A literature search was performed in July of 2014 to retrieve publications related to pain documentation that were published before August of 2014. Multiple electronic databases such as Pub Med, Google Scholar, Google, and the Agency for Healthcare Research and Quality (AHRQ) National Guideline Clearinghouse (NCG) website was searched using the keywords "pain" and "documentation" or "chronic pain and documentation" or "acute pain and documentation" or "pain" and "guidelines" or "geriatric pain" or "pediatric pain". Our inclusion criteria were peer-reviewed publications related to acute or chronic pain documentation. We excluded articles that did not discuss discrete data elements related to pain documentation. In addition, articles that discussed pharmacological management of pain or specific interventions to decrease pain were also excluded. The initial search of the keyword "pain management" or "acute pain management" or "chronic pain management" yielded over 75,000 results. When the additional keyword of "documentation" was added, the results decreased to 455 results. After applying the exclusion criteria of pharmacological management or specific interventions, a total of 50 articles were identified. After reviewing the abstracts to determine if they met inclusion and exclusion criteria and removing overlapping articles, a total of 9 articles were selected for full-text review. After reviewing the content of several articles, it became evident that the Agency for Healthcare Research and Quality website needed to be searched. After searching the website using the keywords "pain guidelines", another 3 web pages within the National Guidelines Warehouse (NCG) powered by the AHRQ were identified. A total of 12 articles and web pages were identified to receive full review.

Categorization of literature and findings

Articles were then categorized into two areas of interest: (1) national and international guidelines on general documentation and treatment, and (2) documentation of pain in certain situations. (e.g. treatment of cancer pain.) After reviewing these articles, new categories were derived based upon the type of pain that was being documented. Pain can be divided in to two major sub-topics: "Acute Pain" and "Chronic Pain."

Define a Comprehensive Reference Model for Pain Documentation that includes Acute and Chronic Pain

After reviewing which data elements were required to be documented, a reference model was created. Then each element definition was compared against the existing reference model that had been derived from HL7, Intermountain and LOINC. The outcomes of each data element were placed into three categories (1) Valid (V), (2) Missing Required (MR), (3) Missing Not Required (MN). The criteria for the data elements that were missing and required were elements that were present in the reference model derived from published sources and not present in the existing reference model. Criteria for the elements that were considered missing but not required were elements that appeared in the existing reference model and not in the reference model derived from published sources. Validated elements were considered a 1:1 match or their definitions between each reference model were similar.

Results

Comprehensive Reference Model for Pain Documentation

Sub-categories of acute pain and chronic pain were assigned to specific situations if the guidelines stated that certain documentation was required. The sub-categories for acute pain were (1) self-reported pain perceptions following surgery, (2) geriatric pain management and (3) pediatric pain management. The sub categories for chronic pain were (1) malignant pain and non-malignant pain. Additional sub-types for non-malignant pain needed to be identified to which were (1) neuropathic pain, (2) musculoskeletal pain and (3) inflammatory pain.

Acute Pain

Acute pain has a set of data elements that make up its reference model. (Table 1). Acute pain consists of three sub-domains, (1)self-reported pain perceptions following surgery,

Table 2 – Self- Reported Pain Reference Model

Reference Model Elements[1]	Elements derived from published sources[1]	Validation Outcome[2]
Pain Onset	Pain History	*V*
Pain Onset (Hours ago)	Pain History	*V*
Speed of Pain Onset	Pain History	*V*
Pain Primary Location	Location	*V*
Pain Quality (Character)	Quality of Pain	*V*
Pain Periodicity	Frequency	*V*
Pain Temporal Pattern	M	*MN*
Pain Alleviating Factors	Alleviating Factors	*V*
Pain Aggravating Factors	Aggravating Factors	*V*
	Pain Duration	
Pain Duration (Hours, Minutes)		*MN*
Associated Signs and Symptoms	Physiological and behavioral Responses to Pain	*V*
Patient Severity Score	Intensity of Pain	*V*
Pain Course	Pain Treatment	*V*
Relative Temporal Context	Quality of Life	*V*
M	Patient Education regarding pain management	*MR*

1 M= Missing
2 V= Validated, MN= Missing and Not Required, MR= Missing and Required

(2) geriatric pain management and (3) pediatric pain management.

These three sub-domains were chosen because of the specificity of how pain needs to be reported. The geriatric and pediatric pain populations were singled out due to the way the patients report and respond to pain stimuli. For example, cognitive status in geriatric pain management is significant because the self-reported pain may not be entirely accurate if the patient is confused. (Table 3). Self-reported pain (Table 2) also has its' specific requirements such as patient education regarding pain management while geriatric pain management required a cognitive status assessment as part of its documentation. Pediatric pain requires only Pain History, Pain Quality, Pain Pattern, Alleviating Factors, Aggravating Factors and Pain Intensity to be documented . In Self-reported perception of pain, it was found that the patient education regarding pain management is a required criteria to be documented. All other elements were considered to either be valid or missing but not required.

Table 3- Acute Pain: Geriatric Pain Management

Reference Model Elements[1]	Elements derived from published sources[1]	Validation Outcome[2]
Pain Onset	Present Pain	*V*
Pain Onset (Hours ago)	Pain History	*V*
Speed of Pain Onset	Pain History	*V*
Pain Primary Location	Pain Location	*V*
Pain Quality (Character)	Pain Character	*V*
Pain Periodicity	Frequency	*V*
Pain Temporal Pattern	Pattern	*V*
Pain Alleviating Factors	Alleviating Factors	*V*
Pain Aggravating Factors	Precipitating Factors	*V*
	Pain Duration	
Pain Duration (Hours, Minutes)		*V*
Associated Signs and Symptoms	Cognitive Status, mental state and functional status	*V*
Patient Severity Score	Pain Intensity	*V*
Pain Course	Pain History	*V*
Relative Temporal Context	Quality of Life	*V*

1 M= Missing
2 V= Validated, MN= Missing and Not Required, MR= Missing and Required

Chronic Pain

Chronic pain consists of two major sub-types: cancer-related pain and non-malignant pain (Table 4). Unlike acute pain, chronic pain sub-domains within the sub-types. Non-malignant pain is broken down in to 3 sub-types: Musculoskeletal and inflammatory pain, and neuropathic pain (Table 5). The minor differences between Data elements can be applied to chronic pain documentation using the Institute for Clinical Systems Improvement (ICSI) Assessment and Management of Chronic Pain Algorithm. [10] The reference models were found to be valid with no elements that were missing and required to be documented upon.

Discussion

Despite the fact that many of the data elements that are necessary for pain documentation are similar, the context of which they are being used can be different. A patient with acute pain is focused on how the pain began and how to resolve it. Using the definitions of the data elements within the sub-domains as a guideline that assist in resolving the problem because the documentation is very specific to each specialty and the sub-domains address the specifics in very broad terms with the understanding that they align to a specific data element within the current reference model. (e.g., Acute abdominal pain for GI specialty versus chronic back pain found in Orthopedics.) Validating each element in the reference model allows a one-to- many ratio to meet the needs in each situation.

Acute pain patients can convert to chronic pain after the initial injury has resolved. In this case, the patient would also convert over to the chronic pain sub-domain. The documentation would then qualify for a specific sub type and secondary sub-domain based upon the signs and symptoms. The implication for the reference model is that it can be used in different scenerios without having to be significantly changed. However, when considering the context of self-reported acute pain, it is necessary to understand the level of education the patient has regarding his or her pain management. In this case, it would be necessary to add the data element "Patient education regarding patient management" to this context in order to gain an accurate picture of the patient's pain and how it is being managed.

Chronic pain is focused on management of the pain overtime, as it may never resolve. By definition, chronic pain is pain persisting longer than three to six months, beyond the time that healing normally occurs.[11] Changes in the chronic pain pathways are usually permanent, may be present in the absence of an identifiable source, and have varying response to conventional analgesic medications.[5]

Table 4- Chronic Pain: Malignant and Non-Malignant

Reference Model Elements[1]	Elements derived from published sources[1]	Validation Outcome[2]
Pain Onset	M	*MN*
Pain Onset (Hours ago)	M	*MN*
Speed of Pain Onset	M	*MN*
Pain Primary Location	Pain Location	*V*
Pain Quality (Character)	Pain Quality	*V*
Pain Periodicity	M	*MN*
Pain Temporal Pattern	Pain Pattern	*V*
Pain Alleviating Factors	Pain Relief	*V*
Pain Aggravating Factors	Mechanism of Pain	*V*
Pain Duration (Hours, Minutes)	Pain Duration	*V*
Associated Signs and Symptoms	M	
Patient Severity Score	Pain Intensity	*V*
Pain Course	M	*MN*
Relative Temporal Context	Functional Ability	*V*
Patient Stated Goal	Follow-Up Plan	*V*

1 M= Missing
2 V= Validated, MN= Missing and Not Required, MR= Missing and Required

The definition of these data elements is one step in the process of configuring clinical documentation content in an EHR. The configuring of how the data elements are represented on a form is a separate and significant and complex process requiring information about the technical capabilites of the system and end-user requirements.

Limitations

The definitions were derived from national guidelines which could be in the process of being updated and thus making these definitions out of date at some point in the future. However, timely integration of the most up to date literature is a continuous struggle for all evidence-based practice and documentation. The search parameters were using keyword searches which, if not using the correct keyword, could have left a result out that may have made a significant contribution to the meaning of the definitions.

Conclusion

Validating a reference model using documentation derived from published sources is only a first step in creating standardized documentation across the entire healthcare system. The method of using evidence to validate reference models for specific clinical topics will be applied to other areas such as wound care, line placement, and living situation. Optimization of clinical data definitions within an EHR necessitates the assessment of broad clinical topics and validation of reference models for each topic that can be applied across settings and specialties in order to create consistent data output.

Acknowledgements

Thank you to the Partners eCare Smart Data Element Workgroup for all their efforts and contributions in this project.

Table 5- Non-Maglinant Chronic Pain: Neuropathic Pain

Reference Model Elements[1]	Elements derived from published sources[1]	Validation Outcome[2]
Pain Onset	*M*	*MN*
Pain Onset (Hours ago)	*M*	*MN*
Speed of Pain Onset	*M*	*MN*
Pain Primary Location	Pain Location	*V*
Pain Quality (Character)	Pain Quality	*V*
Pain Periodicity	*M*	*MN*
Pain Temporal Pattern	*M*	*MN*
Pain Alleviating Factors	Pain Relief	*V*
P ain Aggravating Factors	*M*	*MN*
Pain Duration (Hours, Minutes)	Pain Duration	*V*
Associated Signs and Symptoms	Mental Status, Psychological/Social Factors	*V*
Patient Severity Score	Pain Intensity	*V*
Pain Course	*M*	*MN*
Relative Temporal Context	Functional Ability	*V*
Patient Stated Goal	Goals	*V*

1 M= Missing
2 V= Validated, MN= Missing and Not Required,
MR= Missing and Required

References

[1] Min H, Ohira R, Collins MA, et al. Sharing behavioral data through a grid infrastructure using data standards. *J Am Med INFORMATICS Assoc*. 2014;21:642-649. doi:10.1136/amiajnl-2013-001763.

[2] Goossen WTF. Detailed clinical models: representing knowledge, data and semantics in healthcare information technology. *Healthc Inform Res*. 2014;20(3):163-172. doi:10.4258/hir.2014.20.3.163.

[3] Purser L, Warfield K, Richardson C. Making pain visible: an audit and review of documentation to improve the use of pain assessment by implementing pain as the fifth vital sign. *Pain Manag Nurs*. 2014;15(1):137-142. doi:10.1016/j.pmn.2012.07.007.

[4] Burke CE, Piper J, Calderón J. Inconsistent documentation of drug-related allergies and adverse effects in patient charts. *Am J Heal Pharm AJHP Off J Am Soc Heal Pharm*. 1998;55(3):289-290. http://ezproxy.neu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=9492262&site=ehost-live&scope=site.

[5] DeRemer CE, Fleming VH, Brown S, May JR. Evaluation of pain management documentation. *South Med J*. 2011;104(9):629-633. doi:10.1097/SMJ.0b013e3182296f20.

[6] Health Level Seven International - Homepage. http://www.hl7.org/. Accessed December 22, 2014.

[7] Intermountain Clinical Element Model Browser. http://www.clinicalelement.com/partials/user-guide.html. Accessed December 22, 2014.

[8] Logical Observation Identifiers Names and Codes (LOINC®) — LOINC. http://loinc.org/. Accessed December 22, 2014.

[9] Innovation and Leadership | Partners eCare | Partners HealthCare. http://www.partners.org/Innovation-And-Leadership/Partners-eCare.aspx. Accessed December 22, 2014.

[10] Hooten WM, Timming R, Belgrade M, Gaul J, Goertz M, Haake B, Myers C, Noonan MP, Owens J, Saeger L, Schweim K, Shteyman G WN. Assessment and Management of Chronic Pain Health Care Guideline : Assessment and Management of Chronic Pain. *Inst Clin Syst Improv*. 2013;(November). https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_neurological_guidelines/pain/.

[11] Guideline HC. Assessment and Management of Chronic Pain Health Care Guideline : Assessment and Management of Chronic Pain. 2013;(November).

[12] Saigh O, Triola MM, Link R. Brief report: Failure of an electronic medical record tool to improve pain assessment documentation. *J Gen Intern Med*. 2006;21:185-188. doi:10.1111/j.1525-1497.2006.00330.x.

Address for Correspondence

Partners eCare
2 Ave de Lafayette Ste. 1
Boston, MA 02111
Emily Gesner DNP, RN-BC
(617) 857-5307
egesner@partners.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-810

Consumer Health Information Needs and Question Classification: Analysis of Hypertension Related Questions Asked by Consumers on a Chinese Health Website

Haihong Guo[a], Jiao Li[a], Tao Dai[a]

[a] *Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China*

Abstract

This study built up a classification schema of consumer health questions which consisted of 48 quaternary categories and 35 annotation rules. Using such a schema, we manually classified 2,000 questions randomly selected from nearly 100 thousand hypertension-related messages posted by consumers on a Chinese health website to analyze the information needs of health consumers. The results showed questions in the categories of treatment, diagnosis, healthy lifestyle, management, epidemiology, and health provider choosing were 48.1%, 23.8%, 11.9%, 5.2%, 9.0%, and 1.9% respectively. The comparison of the questions asked by consumers and physicians showed that their health information needs were significantly different (P<0.0001).

Keywords:
Health Consumers; Information Needs; Hypertension; Question Classification; Consumer Health Informatics.

Introduction

The Internet is increasingly becoming a main resource for consumers to acquire health information. 80% of internet users (i.e. 59% of all adults) in the U.S. have looked online for health information [1]. In China, health channels of main portals and professional health websites have become one of the main resources for health consumers [2], and the number of internet users has increased to about 632 million in 2014 [3]. Many researches have proven that health related information online could impact consumers' health decisions and behaviors [1,4]. Health websites that target specific information needs are burgeoning in response to the high demand and impact [5]. However, making health information available does not necessarily guarantee its accessibility and usability [6]. Consumers have difficulty in expressing their information needs using accurate medical query terms, and further failed to retrieve relevant health information [78]. Thus, it is crucial to develop a method to identify health information needs from questions asked by consumers.

In previous studies, a series of templates were developed to guide question composition, and identify meaningful concepts and their relationships, which were further used to construct query strategy [9]. This method requires consumers to follow the scheme when asking questions, which would affect its usability. Some studies used metathesaurus (such as UMLS and the on-going Consumer Health Vocabulary) [10] to assign multi topics of questions so as to construct navigational exploration interface or generate semantic query strategy [11]. It is hard to specify the information needs (concerned aspect), although the topic has been identified. For example, though it could identify that a question was about hypertension, it was difficult to distinguish whether the user wanted to know its diagnosis, treatment, prognosis, or diet, for eaxmple. A promising way to overcome the shortage is to build up a question classification schema to define question topic and its concerned aspects [12].

For professional health-related questions, several classification schemas have been developed, such as the International Classification of Primary Care [13] and the Taxonomies of Generic Clinical Questions (TGCQ) [14,15]. They have proven to be useful when analyzing physicians' and case managers' information needs [15,16] but not suitable for consumers' questions due to the diffference of information needs between physicians and consumers [17,18]. A Layered Model of Context for Consumer Health Information Searching (LMCC) intended to describe consumers' interest topics on cognitive layers [19] but not define the classficassion schema in a systematic manner. In this study, we aim to build up a consumer health question classification schema to understand and specify users' informational needs.

To test the usefulness of classification schema, we used hypertension related questions asked by Chinese consumers as a test dataset. Hypertension has become the main risk factor of over half of cardiovascular diseases, such as stroke and coronary heart disease. There were 270 million patients (that is, at least 2 patients out of 10 alduts) with hypertension in China in 2012, and the number has continued to increase at a rate of 3.1% per year [20]. Thus, hypertension-related questions have become more frequently asked with large variability on the Internet.

Materials and Methods

Data Collection

Messages posted by health consumers from 01 January to 10 August 2014 with a tag of "hypertension (高血压)" or "blood pressure (血压)" under the Q&A (有问必答) section on xywy.com (http://www. xywy.com/, a Chinese health website with over 35 million users) were collected and imported into a MySQL database. The resulting database included 98,032 messages. To conduct an in-depth analysis of the questions contained in the messages, we randomly selected 2,000 messages.

"Question" is defined as a request that a consumer posted to the website on a certain subject to seek answers from professionals, which was identified based on meaning, not form. This study was focused on questions related to

hypertension (高血压), which was sometimes expressed as "high blood pressure (高血压)," or simply as "high pressure (高压)." So, we manually discarded messages that did not match the definition and that were irrelevant to hypertension, but with the similar words such as "high pressure oxygen (高压氧)," "hyperbaric cabin (高压舱)," "high voltage(高压电)," "pressure cooker (高压锅)," etc.. Another new message was randomly selected from the database when an irrelevant message was discarded from the sample, so as to keep the sample size at 2,000.

Classification Schema Development

A topic-based classification schema was developed based on TGCQ [14,15] and LMCC [19]. The categories of clinical related questions (including diagnosis, treatment, management, epidemiology, and their narrower terms) were mainly selected from TGCQ, while the categories of non-clinical questions (such as healthy lifestyle and health provider choosing, and their narrower terms) were mainly selected and expanded from LMCC. We divided some categories into more specific sub-categories so as to code the specific information needs. For instance, the diet category, under healthy lifestyle, was further divided into five tertiary categories, including how to eat, food choosing, interactions, action mechanism, and general.

One of the authors (specialized in medical informatics), classified all the 2,000 sample questions following the classification schema. During the manual classification progress, some categories were added to accommodate questions that did not fall into any existing medical or non-medical specialty. We developed a list of annotation rules and enumerated some general question types for each of the smallest category, so as to improve consistency among coders and the usability of the classification. For example, questions as the following types were coded as 1.1.1.1 (diagnosis→interpretation of clinical finding→symptom):

- I have (or somebody else has) symptom x, what's the condition / matter? (我有（某人有）症状X，是什么情况？ / 是怎么回事？)
- What cased symptom x? (是什么引起症状X？)
- What's the cause of symptom x? (症状X的原因是什么？)

The website (xywy.com) provides a template for users to generate question including three parts: (1) *describe your health status (病情描述)*, (2) *treatments or tests in the past (曾经的治疗或检查情况)*, and (3) *what kinds of help do you want (想得到怎样的帮助).* This template might lead to consumers' confusion on question fill-up. To deal with this case, we developed a rule: if there is the phrase "what kinds of help do you want (想得到怎样的帮助)" in the message, then code the first question after the phrase as the main topic, and successively code the questions followed by as minor topics. Otherwise, we coded the first question in the message as the main topic.

In this way, we developed the preliminary classification schema of consumer health questions, which had 101 topic categories and 32 annotation rules. The classification schema and questions coding was then modified by the following steps:

Firstly, four volunteers (two with medical education backgrouds, the others with informatics backgroud) used the classification to independently code 200 questions randomly selected from the sample, and each volunter made suggestions to specify the rules and increase some categories to accommodate the questions. The author compared the consistency of the five coding results (including the result of herself), and categorised the 200 questions into three groups: all annotators agreed on the topic (n=73), only one annotator disagreed (n=63), and more than one annotator disagreed (n=64). Then we focused on the last group, looking for problematic and ambiguous questions. Analysis of these inconsistencies allowed us to address ambiguous elements in the classification via specifying annotation rules and changing the description of the example general question types.

Secondly, the revised classification was distributed to the five annotators who independently annotated another 300 questions randomly selected from the remaining 1,800 samples. This step was done to measure the interrater reliability of the revised classification schema, as well as to modify it further.

Lastly, the three volunteers annotated the remaining 1,500 messages. Each of them annotaed 500 independently to ensure all of the 2,000 sample messages were annotated by at least two annotators. The codes agreed upon by this step were regarded as the final ones. Then we calculated the number of questions in each topic category, and deleted categories that did not have any questions filled in (e.g. physical characteristics of drugs, pharmacodynamics, mechanism of drug action).

Statistical Analysis

Descriptive analysis was used to calculate the frequency of question topics (main topic only, and all topics respectively). The kappa=$(P_o-P_e)/(1-P_e)$ statistic, which could correct agreement that occured by chance, was used to determine the interrater reliability of the question classification, where P_o was the observed agreement and P_e was the agreement expected by chance. When the number of categories was large, as in this study, P_e would be close to zero, and the kappa value would be close to P_o. Thus we directly used P_o as kappa value. The bigger the kappa value, the better the agreement. We supposed that when the user asked more than one question, it was acceptable to answer any one of them. Therefore, a liberal reliability criteria was used; a match was recorded if either the main or minor topics assigned by one annotator matched the other's assignment.

We merged the topic classification of consumer health questions developed in this study and the Taxonomy of Generic Clinical Questions (TGCQ) [15] into one classification table, and then used a chi-square test to compare the frequency distributions of topics asked by consumers and professionals.

Results

Classification of Consumer Health Questions

The final classification schema was a four hierarchical levels of specificity, consisted of 48 quaternary categories (Table 1), and included 35 annotation rules, down from 101 categories in the preliminary version. The first level included seven broad areas: diagnosis, treatment, management, epidemiology, healthy lifestyle, health provider choosing, and other. Condition/finding management questions asked what steps to take without distinguishing between diagnostic steps and

therapeutic steps [15]. To answer them, one should first give a diagnosis, and then the suggestion of treatment. A branching structure of secondary, tertiary, and quaternary levels describes more and more specific topics of the questions. One or more closely related generic questions were listed for each quaternary category. For instance, the question "A 65-year-old man with unsteady high blood pressure, what's the best blood pressure drug to eat? (65岁老人血压高经常不稳定，吃哪种降压药最好?)" would be coded as 2.1.2.1 (treatment→drug therapy→efficacy/ indications→treatment), and the generic question type could be "Condition y, what's the best drug (to eat / take / use)?"

Table 1 – Classification of consumer health questions.

Code	Primary	Secondary	Tertiary	Quaternary	Frequency(%) (main code)	Frequency(%) (all codes)
1.1.1.1	diagnosis	interpretation of symptom			39(2.0)	44(1.7)
1.1.2.1		clinical finding	sign		146(7.3)	152(5.8)
1.1.3.1			test finding		7(0.4)	8(0.3)
1.1.4.1			multiple findings		286(14.3)	302(11.6)
1.2.1.1		criteria			35(1.8)	37(1.4)
1.3.1.1		test	indications/ efficacy		36(1.8)	55(2.1)
1.3.2.1			accuracy		4(0.2)	6(0.2)
1.3.3.1			timing		6(0.3)	6(0.2)
1.3.4.1			method		4(0.2)	5(0.2)
1.4.1.1		orientation	condition		4(0.2)	5(0.2)
1.5.1.1		cost			0(0.0)	2(0.1)
2.1.1.1	treatment	drug therapy	how to use	general	4(0.2)	7(0.3)
2.1.1.2				dosage	8(0.4)	11(0.4)
2.1.1.3				timing	38(1.9)	52(2.0)
2.1.2.1			efficacy/ indications	treatment	324(16.2)	389(14.9)
2.1.2.2				prevention	3(0.2)	7(0.3)
2.1.3.1			adverse effects	caused by drug	29(1.5)	46(1.8)
2.1.3.2				control	2(0.1)	5(0.2)
2.1.3.3				safety/contraindications	23(1.2)	27(1.0)
2.1.4.1			interactions		20(1.0)	25(1.0)
2.1.5.1			name		1(0.1)	2(0.1)
2.1.6.1			cost		1(0.1)	1(0.0)
2.1.7.1			availability		1(0.1)	1(0.0)
2.1.8.1			brand		2(0.1)	3(0.1)
2.2.1.1		not limited to but may include drug therapy	efficacy/ indications	treatment	473(23.7)	621(23.8)
2.2.1.2				prevention	7(0.4)	16(0.6)
2.2.2.1			timing		4(0.2)	8(0.3)
2.2.3.1			how to do it		1(0.1)	1(0.0)
2.2.4.1			safety/ contra/ sequelae		12(0.6)	26(1.0)
2.2.5.1			cost		2(0.1)	8(0.3)
3.1.1.1	management	condition/ finding			114(5.7)	136(5.2)
4.1.1.1	epidemiology	prevalence			0(0.0)	1(0.0)
4.2.1.1		etiology	causation/ association	risk factors	111(5.6)	149(5.7)
4.2.1.2				genetics	3(0.2)	4(0.2)
4.3.1.1		prognosis			51(2.6)	82(3.1)
5.1.1.1	healthy	diet	how to eat		4(0.2)	4(0.2)
5.1.2.1	lifestyle		food choosing	efficacy	69(3.5)	97(3.7)
5.1.2.2				contraindications	29(1.5)	40(1.5)
5.1.3.1			interactions		2(0.1)	4(0.2)
5.1.4.1			general		19(1.0)	32(1.2)
5.2.1.1		exercise			7(0.4)	15(0.6)
5.3.1.1		weight-losing			3(0.2)	3(0.1)
5.4.1.1		mood control			2(0.1)	3(0.1)
5.5.1.1		general			35(1.8)	107(4.1)
6.1.1.1	health provider	hospital			10(0.5)	18(0.7)
6.2.1.1	choosing	department			11(0.6)	25(1.0)
6.3.1.1		doctor			3(0.2)	6(0.2)
7.1.1.1	other				5(0.3)	6(0.2)
Total					2000(100)	2610(100)

General Topics of Questions Asked by Health Consumers

The 2,000 sample messages were coded with 2,000 main codes and 610 minor codes, the frequency of each topic category was shown in Table 1. 48% of the questions were asked about treatment, which indicated that nearly half of the health consumers posting questions on the website have noticed that they or somebody they care about has had some health problem and needed to be treated. Almost half (45.9%) of the treatment questions were referred in particular to drug

therapy, including how to use drugs (5.6%), how to choose drugs for a particular condition (31.5%) and adverse effects of drugs (6.2%). 23.8% of the questions were asked about diagnosis, and the majority (19.4%) were seeking interpretation of consumers' specific clinical findings in reality as each pertained to the symptom they felt (1.7%), the sign they knew from physical examination (5.8%), or multiple kinds of findings they got (11.6%). 5.2% of the questions were coded as 3.1.1.1 (management of condition or findings) because they were not specified in diagnosis or treatment, and more than half of them (54.4%) were just enumerated as a series of clinical findings without any interrogative sentence or term.

11.9% of questions were asking what to do or what not to do in everyday life in order to keep healthy or get well from certain illnesses. More than half (58.4%) of them were concerned with diet or nourishment. Among the 9.0% of epidemiological questions, 5.7% were about risk factors, both risk factors of the diseases they had and if their condition would be harmful to some particular conditions, e.g., pregnancy, parturition and sexual life. 3.1% of questions were about prognosis and we thought many of them were mainly expressing anxiety as the asker wanted to get an affirmative reply to allay their worry [19]. Among the 1.9% provider choosing questions, half were about medical department choosing for specific conditions or clinical findings, which indicates that medical guide service would be a promising area for health websites.

Interrater Reliability

The kappa statistic for the five coders was 0.63 in the quaternary level of the classification, indicating "substantial" reliability. When just the the primary and secondary levels were considered, the kappa value increased to 0.75. When only the seven broad areas in the primary level were considered, agreement was almost perfect (kappa=0.82).

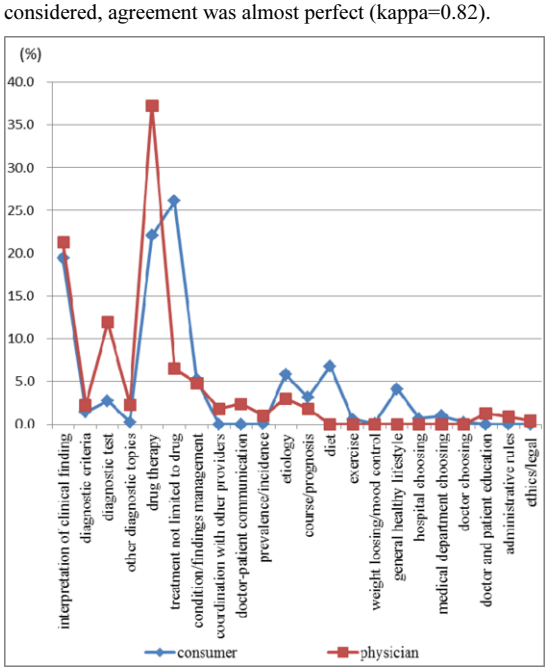


Figure 1 – The frequency distribution of the secondary topics of the questions asked by health consumers and physicians.

Difference of Health Information Needs Between Consumers and Physicians

The chi-square test of the topic distributions of the questions asked by the two groups respectively showed that health information needs of consumers were significantly different from clinical information needs of physicians. The pearson chi-square value on the quaternary category was 1477.89 ($P<0.0001$), and was 854.38 ($P<0.0001$) on the secondary category. Figure 1 shows the frequency difference along the secondary topics of the questions between health consumers and physicians. For exmaple, phyicians are more interested in the diagnosis test than consumers (11.9% vs. 2.8%).

Discussion

Health consumers and physicians both asked questions about diagnosis, treatment, management of conditions and findings, and epidemiology. Besides these topics, health consumers also asked how to keep healthy or help recovery in daily life, because many of them recognized that lifestyle, such as diet, exercise, weight loosing, and mood control, would impact their health status as well [21]. While physicians seldom asked these questions during a patient encounter, it might be because they mainly focused on medical service rather than lifestyle advice [18]. Similarly, health consumers never asked questions about coordination with other providers, doctor-patient communication, doctor and patient education, administrative rules, ethics, and legal issues, since these tasks were usually regarded as health providers' responsibility.

Both of the two groups sought answers for the interpretation of clinical findings, while the questions posted by health consumers were much more vague, the frequency of questions with multiple findings were two times more than that of physician-based inquiries. It might be because they could not distinguish which findings were most important, so they tended to put all the findings they knew. Physicians were more concerned ahout what test to choose for a particular situation and when or how to do it (the question frequency was three times more than that of health consumers), because they wanted to know how to diagnose [14,15], while health consumers wanted to know the diagnosis. Though the frequency of treatment questions was almost equal in the two groups (48.2% vs. 43.7%), physicians' questions were more specified to drug therapy (37.2% vs. 22.1%), and they sometimes asked those questions on very specialized sides, such as composition, pharmacodynamics, action mechanism, and serum levels of drugs, which were rarely asked by health consumers.

Health consumers were mainly concerned about what was wrong with their health (or the health of someone they care about), what went wrong, how to treat it (including choosing which provider to treat), possible adverse drug effects, cross interactions or dangers with other conditions (e.g. pregnancy, breast feeding, etc.), duration of recovery from the illness, and health maintenance in everyday life (e.g. dietary suggestions). Thus, they seldom asked questions that were commonly regarded as the physicians' tasks or too medical specialized.

Conclusion

This study built a classification schema of consumer health questions which consisted of 48 quaternary categories and 35 annotation rules. Five annotators followed this schema and

classified 2,000 questions randomly selected from nearly 100,000 hypertension-related messages posted by consumers on a Chinese health website. The potential uses of this study were identified as follows. First, the classification could be used to organise large collections of consumer health questions, so as to improve retrieval efficiency. Second, the distribution of question topics could be used to guide the building of a knowledge base for health websites, such as setting priorities to building a knowledge base for those frequently asked questions. Third, the coded questions could be used as a corpus for studies, such as training machines to automatically classify the topics of questions posted by health consumers, and further used for the monitoring of hot health topics and automatically generating answers. Last, but not least, the perceived information needs of health consumers could be used to help set the priorities of medical research and patient education.

This study also had some limitations, the sample questions were collected from only one website and defined to be hypertension or blood pressure related; thus, the applicability of the classification on other settings has yet to be studied further. The information needs analyzed in this study were "user based," that is, the topics were assigned without considering the best way to answer it. Although we achieved substantial interrater reliability, surpassed several similar research, such as assigning topics to generic clinical questions (kappa=0.53) [15], and assigning medical subject headings and subheadings (MeSH terms) to journal articles (consistency percentages was 0.43%) [22], the classification and annotation rules have yet to be modified and tested with larger and more diverse sample questions.

Acknowledgments

This research is supported by the National Key Technology Research and Development Program of China (Grant No. 2013BAI06B01). The authors would like to thank Dr. Yueping Sun, Hongyan Liu, Xiaolin He, Ze Zhang, Hui Wang at IMICAMS for their helpful suggestions on data processing.

References

[1] Fox S. The Social Life of Health Information. Washington, DC: Pew Internet & American Life Project. 2011.

[2] Links-group, Health Sohu. 2009 China Health Communication and Popularization Survey. 2009.

[3] China Internet Network Information Center. China Internet Network Development State Statistic Report. 2014.7.

[4] Ying Z. The Causal Relationship Between Health Related Internet Use and Health Behaviors. Wuhan: Huazhong University of Science and Technology. 2011.

[5] Dai T. Enriching online access to health information for the public: efforts from a medical library in China. Chinese Journal of Library and Information Science 2013: 6(2): 1-13.

[6] Zhang Y. A Review of Search Interfaces in Consumer Health Websites. 2011. [Online: https://www.ischool.utexas.edu/~yanz/HCIR2011_Zhang.pdf; Accessed Nov 2014].

[7] Gunther E, and Christian K. How Do Consumers Search for and Appraise Health Information on the World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, and in-Depth Interviews. British Medical Journal 2002: 324(7337): 573-577.

[8] Zeng QT, Crowell J, Plovnick RM, et al. Assisting Consumer Health Information Retrieval with Query Recommendations. Journal of the American Medical Informatics Association 2006: 13(1): 80–90.

[9] Hallett C, Power R, and Scott D. Composing Questions through Conceptual Authoring. Computational Linguistics 2007: 33(1): 105-133.

[10] Qing TZ, and Tony T. Open Source and Collaborative Development of Consumer Health Vocabulary. [online: http://consumerhealthvocab.org/; accessed Nov 2014].

[11] Licong C, Rong X, Zhihui L, et al. Multi-topic Assignment for Exploratory Navigation of Consumer Health Information in NetWellness using formal concept analysis. BMC Medical Informatics and Decision Making 2014: 14: 63-75.

[12] Cao YG, Cimino JJ, Ely J, et al. Automatically Extracting Information Needs from Complex Clinical Questions. Journal of Biomedical Informatics 2010: 43(6): 962-971.

[13] Gebel RS, and Okkes IM. International Classification of Primary Care (ICPC-2-NL). Utrecht: Nederlands Huisartsen Genootschap, 2000.

[14] Ely JW, Osheroff JA, Ebell MH, et al. Analysis of Questions Asked by Family Doctors Regarding Patient Care. BMJ 1999: 319(7206): 358-61.

[15] Ely JW, Osheroff JA, Gorman PN, et al. A Taxonomy of Generic Clinical Questions: Classification Study. BMJ 2000: 321(7258): 429-32.

[16] Schnall R, Cimino JJ, Currie LM, and Bakken S. Information needs of case managers caring for persons living with HIV. J Am Med Inform Assoc 2011: 18(3): 305-8.

[17] Cécile RLB, and Frans JM. Classifying Health Questions Asked by the Public Using the ICPC-2 Classification and a Taxonomy of Generic Clinical Questions: An Empirical Exploration of the Feasibility. Health Communication 2010: 25(2): 175-181.

[18] Reeder B, Le T, Thompson HJ, et al. Comparing Information Needs of Health Care Providers and Older Adults: Findings from a Wellness Study. MedInfo2013: Proceedings of the 14th World Congress on Medical and Health Informatics, 2013, 192: 18-22.

[19] Yan Z. Toward a Layered Model of Context for Health Information Searching: An Analysis of Consumer-Generated Questions. Journal of the American Society for Information Science and Technology 2013: 64(6): 1158-1172.

[20] National Center for Cardiovascular Diseases, China. Report on Cardiovascular Diseases in China (2013). Beijing: Encyclopedia of China Publishing House, 2014.4.

[21] Publicity Department of National Health and Family Planning Commission of the People's Republic of China, Chinese Health Education Center. 2012 Chinese Residents Health Literacy Monitoring Report. 2013:6-12.

[22] Funk ME, and Reid CA. Indexing consistency in MEDLINE. Bull Med Libr Assoc 1983: 71: 176-83.

Address for correspondence

Tao Dai, Institute of Medical Information & Library, Chinese Academy of Medical Sciences, 3rd Yabao Road, Chaoyang District, Beijing 100020, China.

Email: dai.tao@imicams.ac.cn

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-815

Generating and Executing Complex Natural Language Queries across Linked Data

Thierry Hamon[a], Fleur Mougin[b], Natalia Grabar[c]

[a] LIMSI-CNRS, Orsay, France; Université Paris 13, Sorbonne Paris Cité, France
[b] Université Bordeaux, ISPED, Centre INSERM U897, ERIAS, France
[c] STL UMR8163 CNRS, Université Lille 3, France

Abstract

With the recent and intensive research in the biomedical area, the knowledge accumulated is disseminated through various knowledge bases. Links between these knowledge bases are needed in order to use them jointly. Linked Data, SPARQL language, and interfaces in Natural Language question-answering provide interesting solutions for querying such knowledge bases. We propose a method for translating natural language questions in SPARQL queries. We use Natural Language Processing tools, semantic resources, and the RDF triples description. The method is designed on 50 questions over 3 biomedical knowledge bases, and evaluated on 27 questions. It achieves 0.78 F-measure on the test set. The method for translating natural language questions into SPARQL queries is implemented as Perl module available at http://search.cpan.org/~thhamon/RDF-NLP-SPARQLQuery

Keywords:

Natural Language Processing; SPARQL; Linked Data; biomedical domain; semantic resources; frames; Knowledge Databases.

Introduction

The knowledge bases (KBs) that record existing biomedical knowledge are becoming increasingly available, although they are spread over different life-science KBs that usually focus on a specific data type: chemical, pharmacological, and target information on drugs in Drugbank [1]; clinical studies [2]; drugs and their side effects in Sider [3]; etc. Creation of connections between these knowledge bases is an important research area [4]. This is indeed crucial for obtaining a more comprehensive view and is also required for producing new knowledge from the already available data. The knowledge encoded in the KBs and dataset interlinks are usually represented as RDF triples, on the basis of which the linked data can be queried through a SPARQL end-point. However, it remains difficult for typical users of such knowledge sources (mainly physicians and life-science researchers) to manage the syntactic and semantic specificities of the SPARQL language together with the structure of various KBs. In order to be able to efficiently use such KBs it is necessary to design friendly interfaces that mediate the technical and semantic complexity of the task, and to provide simpler approaches for querying the KBs. It has been shown that natural language interfaces may provide the best solution [5]. The main challenge is then to propose efficient methodologies for an easy and reproducible rewriting of natural language questions in SPARQL queries.

We present a method for transforming natural language questions into SPARQL queries, and their application to biomedical Linked Data. The method is based on the use of Natural Language Processing (NLP) methods and semantic resources for enriching questions with specific annotations. Questions are then translated in SPARQL with a rule-based approach. For designing the proposed approach, we use questions proposed by the QALD challenge task 2 [6] and additional questions for evaluation. We work with three KBs (Drugbank, Diseasome, and Sider) and can address questions related to, for instance, drug composition and properties, the related disorders, and adverse effects of drugs.

We first present the related work. We describe the proposed method and semantic resources available and developed for processing and translating the questions. We then present the results obtained and their evaluation.

Related work

It is possible to distinguish two kinds of objectives: design methods that transform natural language questions in SPARQL queries [7-8], and design methods that perform this transformation and also execute the queries over Linked Data. In both cases it is necessary to define the user interface that hides the underlying structure of the knowledge bases and the SPARQL syntax. We investigate the second purpose, which also provides the possibility to properly evaluate the methods and the results obtained. Three possibilities are usually explored for querying the Linked Data [8]: Knowledge-Based Specific Interface, Graphical Query Builder, and Question-Answering System. It has also been demonstrated that, for this task, the use of natural language is preferred to the use of keywords, menus, or graphs [6]. Among the existing work, we can mention the use of Question-Answering systems on 50 questions from DBpedia proposed by the QALD-2 challenge, that gives 0.62 average F-measure obtained with 39 questions. Average recall for these 39 questions is 0.63 and average precision is 0.61, although 11 questions cannot be covered by the templates [9]. Studies exploring KBs that describe general language are frequent [10-11], although there are some studies on processing biomedical KBs. For instance, experiments have been proposed on translation of medical questions from a journal into SPARQL queries. This work combined the SVM machine learning-based approach with patterns to generate the SPARQL queries. The evaluation was carried out on 100 questions and the corresponding queries were tested on clinical documents. The method achieved 0.62 precision [12].

Materials

Questions and reference data are provided by the Question Answering over Linked Data (QALD) challenge 2014 [13], dedicated to retrieval of biomedical information in linked KBs with questions in natural language. This question set is composed of 50 questions which we use for setting up the methods. We create 27 additional questions for testing the method. Examples of the questions processed include:

- Which foods does fluvoxamine interact with?
- Are there drugs that target the Probable arabinosyltransferase A?
- Which genes are associated with subtypes of rheumatoid arthritis?
- Which disease has the highest degree?
- Which targets are involved in immune function?

Our method relies on existing biomedical resources with semantic entities data, and on additional resources collected and built for supporting the method. The steps of the method are described in the next section.

Domain-Specific Resources. We use three biomedical resources to process the question set: Drugbank, Diseasome, and Sider. Drugbank http://www.drugbank.ca is dedicated to drugs [1]. It merges chemical, pharmacological, and pharmaceutical information from other available biomedical KBs. We exploit the documentation [14] of this resource to define the rewriting rules and regular expressions for the named entity recognition. Diseasome [15] is dedicated to diseases and genes linked among them by known disorder/gene associations [16]. It provides a single framework with all known phenotypes and disease gene associations, indicating the common genetic origin of many diseases. We exploit the RDF triples and documentation of the resource to define the rewriting rules. Sider [17] is dedicated to adverse drug effects [18]. It contains information on marketed medicines and their recorded adverse drug reactions. Information is extracted from public documents and package inserts and provides side effect frequency, drug and side effect classifications, and links to other data such as drug-target relations. The content of each resource is provided in specific format: RDF triples *subject predicate object*. For this reason, we can exploit their RDF schema to define frames.

Additional Resources for Question Annotation. On the basis of the RDF triples, frames are built from the RDF schemas in which the RDF predicate is the frame predicate, and the subject and object of the RDF triples are the frame elements. This also includes the OWL *sameAs* triples. Several types of frame entities are isolated:

- Subject, object, and predicate from triples become semantic entities. They may occur in questions. In this way, the frames are the main resource for rewriting questions in queries;
- The vocabulary specific to questions is also built. It covers, for instance, aggregation and negation operators and types of questions;
- RDF literals, issued from named entity recognizer or term extractor, complete the resources. RDF literals are detected with specifically designed automata that rely on the source KB documentation.

These entities are associated with the expected semantic types, which allows creating the queries and rewriting the RDF triples in the SPARQL queries. In that respect, we can consider internationalized resource identifiers (IRI), strings, common datatypes, or regular expressions when literals are expected. Most of the entities are treated through their semantic types, although some ambiguous entities (e.g., interaction or class) are treated atomically. For these, the rewriting rules are applied contextually to generate the semantic entities corresponding to the frames. During the query construction step, semantic types become variables and are used for connecting the edges of queries.

Methods

The method is rule-based and is composed of five parts: linguistic and semantic annotation of questions, question abstraction, query construction, query execution, and evaluation. We demonstrate the main steps of the method with the question, *Give me drugs in the gaseous state*. The method is based on NLP tools and methods and must respect constraints of the SPARQL syntax.

Linguistic and Semantic Annotation of Questions is performed in several steps: identification of numerical values (corresponding to numbers or solubility index, for instance) with a named entity recognizer; parsing of questions in words; part-of-speech tagging and lemmatization of questions with TreeTagger [19] (Figure 1); and annotation of semantic entities (terms and their associated semantic types) with the TermTagger Perl module [20] and semantic resources like disease names and side effects (Figure 2). In order to complete the coverage of semantic resources, we use the term extractor YaTeA [21-22] for the identification of noun phrases.

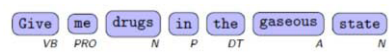


Figure 1 - Linguistic annotation of questions.

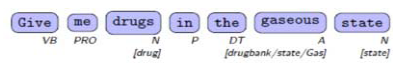


Figure 2 - Semantic annotation of questions.

In the *Question Abstraction* step the objectives are to identify relevant elements within questions and to build a representation of these elements (Figure 3). This step relies on linguistic and semantic annotations. For instance, the word *drug* receives the semantic type *Drugs* from Drugbank. The main difficulty is that some entities may receive conflicting or concurrent semantic types. The purpose is then to select those semantic types that are the most correct and useful for the next steps. This is done as follows:

- We keep larger terms which do not include other semantic entities;
- We define and apply rewriting rules on the training set in order to modify or delete semantic types associated with a given entity according to the context;
- We also modify or delete entities according to their context. For instance, the semantic entity *interaction* must be rewritten in *interactionDrug1* if the mention

of drug is found in its context, but it has to be rewritten in *foodInteraction* if a term with the semantic type *food* is found in its context.

We define a set of 44 contextual rewriting rules based on vocabulary used in questions from the training set and on the documentation of KBs, mainly of Drugbank [14]. After this processing, the identification of relevant elements can be performed more safely. After the identification of relevant elements, we perform the question abstraction, for which we identify information related to the structure of query (i.e., question topic, result form, predicate and arguments, scope of negation, and conjunctions):

- The first entity with the semantic type that is provided by the RDF subjects and objects issued from the resources is considered to be the question topic. This information is used for the query construction. The question topic in Figure 3 is identified as *drug*;

Agregation operator:			
Question topic:	drug (drugbank/drugs)		
Predicates:	Frame		
	drugbank/drugs	state	STRING
Arguments:	Semantic type	Word	SPARQL type
	drugbank/state/Gas	gaseous	STRING/Gas

Figure 3 - Identification of relevant elements in questions.

- For the definition of the result form, the question is scanned in order to identify words expressing negation (e.g. *no*), aggregation operation on the results (e.g. *number* for *count*, *mean* for *avg*, *higher* for *max),* and specific result forms such as boolean queries (*ASK).* At this step the linguistic expressions are associated with the corresponding SPARQL operators. Information extracted at this step is recorded and used for the query construction. In the example in Figure 3, no such information is found;
- Identification of Predicate and Argument: we use linguistic representation of the RDF schema, i.e., frames that contain one predicate and at least two elements with associated semantic types. The potential predicates, subjects, and objects of frames are identified among the semantic entities and recorded: entries are the semantic types of the elements and refer to linguistic, semantic and SPARQL information associated with these elements. Subjects and objects are described with inflected and lemmatized forms of words or terms, the corresponding SPARQL types, and indicators on their use as object or subject of a predicate. Concerning the predicates, only semantic types of their arguments are instantiated. Subjects and objects can be URI, RDF typed literals (numerical values or strings), and extracted terms (these are considered as elements of regular expressions). In Figure 1, the predicate *state* with the expected arguments *drugbank/drugs* and *Gas/String* is recognized.
- Scope of negation and conjunctions: Argument and predicate in the neighborhood of negation and conjunctions are identified. These elements are recorded as negated or coordinated.

Agregation operator:			
Question topic:	?v0		
Predicates:	Frame		
	?v0	state	STRING/Gas
Arguments:	Semantic type	Word	SPARQL type
	drugbank/state/Gas	gaseous	STRING/Gas

Figure 4 - Construction of queries.

The objective of the *Query Construction* step is to connect previously identified elements and to build a semantic representation of the SPARQL graph pattern (introduced by the keyword *WHERE).* Figure 4 presents the architecture of the connection of elements and query construction. Thus, the predicate arguments are instantiated by URIs associated with the subjects, objects, variables, and numerical values or by strings. For each question, we perform several connections:

- The question topic is connected to the predicate(s). A given variable is associated with the question topic and with the predicate argument that matches the semantic type of the question topic. Note that at the end of this step, the question topic may remain unassociated with a predicate. In Figure 4, variable *?v0* represents the association between the question topic and the subject (with the expected type *drugbank/drugs*) of the predicate *state;*
- The predicate arguments are connected to subject and objects identified during the question abstraction: they concern elements referring to URI. Moreover, each predicate within the conjunction scope is duplicated and its arguments are also connected to the subject and objects if needed;
- The predicates are connected between them through their subjects and objects. The connection between two predicates is also represented by a variable;
- The predicates from different datasets are connected. We use the *sameAs* description to identify URIs referring to the same element. New variables are defined to connect two predicates;
- The remaining question topic is connected to arguments of the *sameAs* predicate;
- The arguments corresponding to the *string* type are connected with the extracted terms that are considered as string expressions. They are connected through the string matching operator *REGEX.*

At this point, the predicate arguments which remain unassociated are replaced by new variables in order to avoid empty literals. Finally, the negation operator is processed: the predicates are marked as negated and the arguments within the negation scope are included in a new predicate *rdf:type* if required. At this step, each question is fully translated into a representation of the SPARQL query. Figure 5 illustrates the construction of the query.

```
SELECT DISTINCT ?v0
WHERE {
?v0 <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/state>
"Gas".
}
```

Figure 5 - Construction of queries.

Query Generation is a syntactic step, during which the query is formed syntactically. The SPARQL query representation

built during the query construction step is used to generate the SPARQL query string. The process is composed of two steps:

- Generation of the result form which takes into account the expected type of the result form *(ASK* or *SELECT*), the presence of aggregation operators, and the variable associated to the question topic;
- Generation of the graph pattern. This step consists of generation of strings for representing each RDF triple and defining a SPARQL filter if the predicates are negated. When aggregation operators are used, we also need to recursively generate sub-queries for computing the subsets of expressions, before their aggregation. In example from Figure 1, the predicate *state* is replaced by the corresponding URI and its object is replaced by the string *gas*.

The SPARQL queries have been submitted to a SPARQL end-point. For these experiments, we use the SPARQL end-point provided by the QALD-4 challenge [23] and answers are collected for the evaluation.

Evaluation Metrics. The generated SPARQL queries are evaluated through their answers with three macro-measures [24] shown in Figure 6. *TP(q)* are the correct answers, *FP(q)* are the wrong answers and *FN(q)* are the missing answers for the question *q*. The use of macro-measures equally considers all the questions independently of the number of expected answers to the SPARQL queries. The comparison is done with the answers obtained through the manual querying of KBs.

$$M_{\text{precision}} = \frac{\sum_{i=1}^{|q|} \frac{TP(q)}{TP(q)+FP(q)}}{|q|}$$

$$M_{\text{recall}} = \frac{\sum_{i=1}^{|q|} \frac{TP(q)}{TP(q)+FN(q)}}{|q|}$$

$$M_{\text{F-measure}} = \frac{2 \times M_{\text{precision}} \times M_{\text{recall}}}{M_{\text{precision}} + M_{\text{recall}}}$$

Figure 6 – Macro-measures for evaluating SPARQL queries

Results

Global Results. Table 1 indicates the overall results obtained on the training and test sets. On the test set, the macro-F-measure is 0.78 with 0.81 precision and 0.76 recall while on the training set, the macro-F-measure is 0.86 with 0.84 precision and 0.87 recall. Each question is processed in less than 2 seconds on a standard computer (2.7GHz dual-core CPU and 4 Gb of memory). Most of the computing time is spent for the linguistic and semantic annotation of the questions.

Table 1 - Results obtained with the training and test sets.

Query set	Training (50 Q)	Test (27 Q)
Correct queries	39	20
Micro-precision	0.84	0.81
Micro-recall	0.87	0.76
Micro-F-measure	0.86	0.78

Discussion

Questions from the test set may involve new operators and lexicon not known from the test set. Nevertheless, our method always proposes syntactically correct SPARQL queries for all natural language questions from the training and test sets: 18 questions provide the exact expected answers, two questions return partial answers, and other questions return no correct answers. On the training set, 39 out of 05 SPARQL queries are semantically correct and provide the expected answers, 6 return partial answers, and 5 return no answers. An analysis of erroneous or partial answers shows that most of the errors are due to:

- The encoding of the *sameAs* predicate in the resources. We observe that, although our method generates the correct SPARQL query, the SPARQL end-point does not return the expected answers. Besides, we observe that the correct answers can be obtained by switching the arguments of the *sameAs* predicate in the queries. It appears thus that the instances of this predicate do not encode the expected reflexivity of this relation in the source KBs while our method assumes that the *sameAs* predicate is reflexive by definition;
- The management of ambiguities in questions. The errors due to ambiguities are mainly related to (1) the annotation of semantic entities and (2) the expected meaning of terms in the questions. Related to (1) annotation, in the question, *Which genes are associated with breast cancer?*, *breast cancer* is correctly annotated, while the reference data propose that it is to be associated with the semantic entity *Breast cancer-1*. Related to (2) expected meaning,. semantic entities that occur in some questions may refer to specific entities while in other questions they refer to general entities. For instance, the semantic entity *anemia* in *What are enzymes of drugs used for anemia?* refers to all types of anemia (*Hypercholanemia*, *Hemolytic anemia*, *Aplastic anemia*, etc.) and not to the elements that contain the label *anemia.*

We defined and described some rules for managing the ambiguities, but they must be completed with additional rules. For instance, these two main problems can be solved by using regular expressions in SPARQL graphs rather than URIs. However, we must test the influence of this modification on the whole set of queries. Other erroneous answers happen during the question abstraction step when the question topics are wrongly identified or when the contextual rewriting rules cannot be applied. Errors also occur during the query construction step. The method may abusively connect predicate arguments and semantic entities or, on contrary, it may not consider all the identified semantic entities. Further investigations have to be carried out to solve these limitations.

During the design of queries, we may also have difficulties to express some constraints in SPARQL. For instance, the question, *Which approved drugs interact with calcium supplements?,* requires defining a regular expression with the term *calcium supplement* while this term is only mentioned in conjunction with other supplements (*e.g., Do not take calcium, aluminum, magnesium or iron supplements within 2 hours of taking this medication.*). We assume that solving this difficulty requires more sophisticated NLP processing of the textual elements of the RDF triples: parsing the RDF textual

elements, named entity and term recognition, identification of discontinuous terms and term variants, etc. When a sufficient quantity of questions and corresponding queries is available, some regular expressions used in the current work can be replaced by machine learning approaches in order to make a generalization of the named entity detection, or for the detection of semantic relations and operators.

Other limitations are related to KB updates and changes to their structure. In the former case it is required to rebuild the semantic resources used for identifying the semantic entities. In the latter case the entire frames must be regenerated. This is ongoing research work. Moreover, the possibility of adding new resources, such as Dailymed [25] or RxNorm [26], is a related problem. When managed, this functionality will allow processing new linked datasets with low adaptation cost.

Conclusion

We proposed a rule-based method to translate natural language questions into SPARQL queries. The method relies on linguistic and semantic annotations of questions with NLP methods, semantic resources, and RDF triples description. We designed our approach on 50 biomedical questions proposed by the QALD-4 challenge and tested it on 27 newly created questions. The method achieves good performance with 0.78 F-measure on the set of 27 questions. Future work aims to address the limitations of our current method including the management of term ambiguity, question abstraction, and query construction. Moreover, to avoid the manual definition of the dedicated resources required by our approach (frames, specific vocabulary, and rewriting rules), we plan to investigate how to automatically build such dedicated resources from the RDF schemas of the Linked Data set. It will also facilitate the integration of other biomedical resources such as Dailymed or RxNorm, and the use of our method in text mining applications.

Acknowledgements

This work was partly funded through the project POMELO (PathOlogies, MEdicaments, aLimentatiOn) funded by the MESHS (Maison européenne des sciences de l'homme et de la société) under the framework Projets Emergents.

References

[1] Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J.: Drugbank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research 34, D668–D672 (2006), database issue

[2] ClinicalTrials.gov. Available from:http://clinicaltrials.gov

[3] Kuhn, M., Campillos, M., Letunic, I., Jensen, L.J., Bork, P.: A side effect resource to capture phenotypic effects of drugs. Molecular Systems Biology 6(1) (2010), http://msb.embopress.org/content/6/1/343

[4] Linking Open Drug Data (LODD). Available from: http://www.w3.org/wiki/HCLSIG/LODD

[5] Kaufmann, E., Bernstein, A.: How useful are natural language interfaces to the semantic web for casual end-users? In: Proceedings of the Forth European Semantic Web Conference (ESWC 2007). Innsbruck, Austria (June 2007)

[6] Task 2: Biomedical question answering over interlinked data. Available from: http://greententacle.techfak.uni-bielefeld.de/~cunger/qald/index.php?x=task2&q=4

[7] Pradel, C., Haemmerlé, O., Hernandez, N.: Des patrons modulaires de requêtes SPARQL dans le système SWIP. In: Journées Francophones d'Ingénierie des Connaissances (IC). pp. 412--428 (juin 2012)

[8] Lehmann, J., Bühmann, L.: Autosparql: Let users query your knowledge base. In: Proceedings of the 8th extended semantic web conference on The semantic web: research and applications` Part I, pp. 63--79 (2011)

[9] Unger, C., Bühmann, L., Lehmann, J., Ngomo, A.C.N., Gerber, D., Cimiano, P.: Template-based question answering over RDF data. In: WWW. pp. 639—648 (2012)

[10] Damljanovic, D., Agatonovic, M., Cunningham, H.: Natural language interfaces to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. In: Proceedings of the 7th International Conference on The Semantic Web: Research and Applications - Volume Part I. pp. 106--120. ESWC'10, Springer-Verlag, Berlin, Heidelberg (2010)

[11] Kuchmann-Beauger, N., Aufaure, M.A.: Natural language interfaces for datawarehouses. In: 8èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2012), Bordeaux. RNTI, vol. B-8, pp. 83--92. Hermann, Paris (Juin 2012)

[12] Abacha, A.B., Zweigenbaum, P.: Medical question answering: Translating medical questions into sparql queries. In: ACM SIGHIT International Health Informatics Symposium (IHI 2012) (2012)

[13] Biomedical question answering over interlinked data, http://greententacle.techfak.uni-bielefeld.de/~cunger/qald/index.php?x=task2&q=4

[14] Introduction to DrugBank. Available from: http://www.drugbank.ca/documentation

[15] Diseasome: Explore the human disease network. Available from: http://diseasome.eu

[16] Janjić, V., Pržulj, N.: The core diseasome. Mol Biosyst 8(10), 2614-2625 (Aug 2012), http://dx.doi.org/10.1039/c2mb25230a

[17] http://sideeffects.embl.de

[18] Kuhn, M., Campillos, M., Letunic, I., Jensen, L.J., Bork, P.: A side effect resource to capture phenotypic effects of drugs. Molecular Systems Biology 6(1) (2010), http://msb.embopress.org/content/6/1/343

[19] Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: Jones, D., Somers, H. (eds.) New Methods in Language Processing Studies in Computational Linguistics (1997)

[20] Sider 2 Side Effect Resource. Available from: http://search.cpan.org/~thhamon/Alvis-TermTagger/

[21] Aubin, S., Hamon, T.: Improving term extraction with terminological resources. In: Salakoski, T., Ginter, F., Pyysalo, S., Pahikkala, T. (eds.) Advances in Natural Language Processing (5th International Conference on NLP, FinTAL 2006). pp. 380--387. No. 4139 in LNAI, Springer (August 2006)

[22] Lingua-YaTeA-0.622. Available from: http://search.cpan.org/~thhamon/Lingua-YaTeA/

[23] http://vtentacle.techfak.uni-bielefeld.de:443/sparql (Last accessed November 2014)

[24] Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1--47 (2002)

[25] DailyMed. Available from: http://dailymed.nlm.nih.gov/

[26] National Library of Medicine. RxNorm, a standardized nomenclature for clinical drugs. 2009. Available at www.nlm.nih.gov/research/umls/rxnorm/docs/index.html. Bethesda, Maryland, USA

Address for correspondence

Thierry Hamon, LIMSI-CNRS, Rue John von Neumann, Campus Universitaire d'Orsay, Bât 508, 91405 ORSAY, Cedex, France

thierry.hamon@limsi.fr +33 (0) 1 69 85 80 39

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-821

Analyzing Operative Note Structure in Development of a Section Header Resource

Genevieve B. Melton[a, b], Yan Wang[b], Elliot Arsoniadis[a, b], Serguei V.S. Pakhomov[b,c], Terrence J. Adam[b,c], Mary R. Kwaan[a], David A. Rothenberger[a], Elizabeth S. Chen[d]

[a] *Department of Surgery, University of Minnesota, Minneapolis, MN, USA*
[b] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[c] *College of Pharmacy, University of Minnesota, Minneapolis, MN, USA*
[d] *Center for Clinical and Translational Science, University of Vermont, Burlington, VT, USA*

Abstract

Operative notes contain essential details of surgical procedures and are an important form of clinical documentation. Sections within operative notes segment provide high level note structure. We evaluated the HL7 Implementation Guide for Clinical Document Architecture Release 2.0 Operative Note Draft Standard for Trial Use (HL7-ON DSTU) Release 1 as well as Logical Observation Identifiers Names and Codes (LOINC®) section names on 384 unique section headers from 362,311 operative notes. Overall, HL7-ON DSTU alone and HL7-ON DSTU with LOINC® section headers covered 66% and 79% of sections headers (93% and 98% of header instances), respectively. Section headers contained large numbers of synonyms, formatting variation, and variation of word forms, as well as smaller numbers of compound sections and issues with mismatches in header granularity. Robust operative note section mapping is important for clinical note interoperability and effective use of operative notes by natural language processing systems. The resulting operative note section resource is made publicly available.

Keywords:

Surgical Procedures, Operative; Vocabulary, Controlled; Medical History Taking/methods; Quality and Safety; Documentation; Electronic Health Records

Introduction

Operative notes are traditionally created at the completion of a surgical procedure by the primary surgeon who recalls the procedure details and dictates these into a narrative that is subsequently transcribed. Effective operative note documentation is important for assessing surgical quality (1), billing and medico-legal issues (2, 3), and other secondary uses of operative notes (4). With increasing adoption of electronic health record (EHR) systems, operative notes and other clinical documents are increasingly generated and immediately available. EHR systems also enable other mechanisms for note generation, including voice to text software, typed notes, synoptic reporting, and templated notes.

Synoptic reports are used to create documents with discrete data fields whereby desired information from the note can be collected, stored, and retrieved in a standardized fashion. In contrast, templated notes range in the amount of structure that they contain, including some having highly prescriptive and structured formats to others having mostly free-text narrative with primarily document section structure alone. While dictation and transcription remains the most common mechanism for operative note creation, synoptic reports and templates are increasingly used in surgery for operative report creation and appear to encourage improved completeness of these documents (5, 6). Independent of the mechanism used to create the document, section headers in operative notes provide high level structure and serve as "containers" which provide context to the text within the given section (7).

Previous research has examined section headers in clinical notes, with some work exploring the task of automated classification of sections in documents. As an initiative with the OpenGALEN project, Mori et al. utilized "tags" in 5 clusters (Nature, Safety Context, Interpretation, Intention, and Organization) and evaluated this approach to classify 600 section headings (7). Denny et al. reported on a terminology of document sections in "history and physical" (H&P) notes and developed an associated section terminology with Logical Observation Identifiers Names and Codes (LOINC®) mappings (8). These authors later developed an algorithm, "SecTag", to identify and label section headers and section boundaries in H&P notes. Similarly, others have utilized predefined sections to aid in a number of natural language processing tasks, such as problem list extraction and named entity recognition (9-11).

In the United States, H&P note formats are largely governed by Evaluation and Management (E/M) documentation, which provides guidelines for assessing adequacy of documentation for each patient encounter resulting in a "level of service" and justification for a patient bill (12). Operative notes, in contrast, are not covered by E/M, and the Joint Commission (13) and Centers for Medicare & Medicaid Services (CMS) (14) have specified criteria for operative notes including information on suggested contents and note sections. Overall, Joint Commission designates eleven required elements for operative notes: name(s) of primary surgeon/ physician and assistants, pre-operative diagnosis, post-operative diagnosis, name of the procedure performed, findings of the procedure, specimens removed, estimated blood loss, date and time recorded, indications for the procedure, intra-operative complications, and a full description of the procedure.

The Health Level 7 (HL7) Structured Documents Workgroup seeks to develop structured healthcare document standards to promote document and data interoperability. This group has created implementation guides for clinical documents including one for operative notes, the Implementation Guide for HL7 Clinical Document Architecture (CDA) Release 2.0, Operative Note Draft Standard for Trial Use Release 1 (HL7-ON DSTU) (15). This specification includes Level 1 (header constraints), Level 2 (section level constraints of the structuredBody of the ClinicalDocument), and Level 3 (entry

level constraints within a section; specifying only the Plan section of Operative Notes) requirements. The HL7-ON DSTU was created using a variety of data sources and expert opinions including subject matter expert input, summary statistics from sample operative notes, Joint Commission Operative Note Requirements: Standard IM.6.30 (13), and CMS Operative Note Requirements (14). Where possible, the HL7-ON DSTU utilizes existing clinical statement entries and Continuity of Care Document (CCD) elements, and other Implementation Guide templates. As such, some items considered clinical statement entries in other contexts are treated as sections. The HL7-ON DSTU also maps section headings using LOINC® where available.

To improve available resources and tools for clinical natural language processing[1] specifically for operative reports (16, 17) and using our experience with clinical standards evaluation (18-20), we sought to use the HL7-ON DSTU and LOINC® section codes to represent operative note section headers and to develop a resource for operative note section headers.

Materials and Methods

Study Overview

Figure 1 provides a high level summary of this study. The HL7-ON DSTU was examined and section headers with associated LOINC® codes were collected along with potential document section headers from LOINC®. In this case, LOINC® was used as the clinical terminology for mapping section headers, which are considered an observation or measurement. Although the Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT) contains many clinical concepts, section headers and document names are not represented. All operative notes over a 4-year period from University of Minnesota-affiliated Fairview Health Services, which includes an academic medical center, one children's hospital, four community hospitals, and three ambulatory surgery centers, were collected from a full range of general surgery and surgical subspecialties, and section headers were extracted. Headers were mapped and coded to eliminate non-section headers, assess section header variation, and identify granularity issues with mapping to structured sections. The section headers and mappings were combined into a resource. Institutional review board approval was obtained and informed consent waived for this minimal risk study.

HL7-ON DSTU Section Header Extraction

Level 1 and Level 2 HL7-ON DSTU constraints were used in this evaluation. Level 1 header elements included information commonly contained in sections for operative notes (e.g., "Primary Performer" - typically referred to as the "Surgeon" or "Primary Surgeon" in operative notes is a header element). Required and optional operative note section names were used along with LOINC® section mappings, section descriptions and suggested information about each item. Level 3 constraints were excluded from the analysis as were Level 1 header elements not related to section headers (such as elements to encode the overall operative note specification).

In addition to the HL7-ON DSTU section headers, LOINC® section header names, codes, and descriptions were collected by extracting entries of "DOCUMENT_SECTION", resulting in 121 distinct sections from LOINC® Version 2.42. Unmapped terms with this list were also mapped to "CLASS" entries of "H&P.HX", "H&P.HX.LAB", "H&P.PX", or "H&P.SURG PROC" with free text search in a second step.

[1] http://healthinformatics.umn.edu/research/nlpie-group

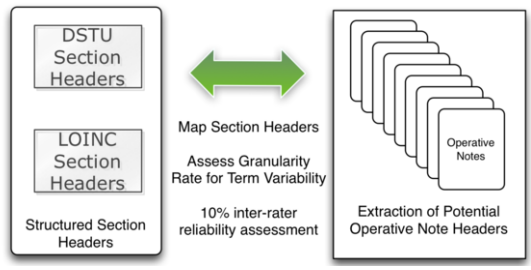


Figure 1 – Overview of Study

Operative Note Section Header Evaluation

Potential operative note section headers were automatically extracted using heuristic rules including the use of capitalization, semi-colons and hyphens, and line-breaks. Frequencies of each potential header were calculated and a cut-off of 100 was used in coding headers. The eliminated set of headers accounted for less than 2% of overall entries.

Each header was then manually designated as a document title, a potential section header, or not a potential section header by two coders. Potential section headers were then mapped to a HL7-ON DSTU standard section name and the provided LOINC® section specification, if available. Table 1 provides an overview and examples of section header codings.

Table 1 – Coding for Operative Note Section Headers

Coding and Explanation	Example(s)
Not a Section Header: *term is not a known section header or document title*	"See Radiology Report From"; "Operating Room"
Document Title: *term is a document title*	"Brief Operative Note"; "Operative Report"
Document Header Information: *term is other document information*	"Patient Identification"; "Dept"
Correct Section Header: *term is preferred section in* HL7-ON DSTU	"Anesthesia"; "Complications"; "Surgery Description"
New Section Header: *term not in* HL7-ON DSTU	"Cross-Clamp Time"; "Preoperative History"
Synonym: *term is alternate synonymous section name (new or known section)*	"Surgery Description" (*Preferred)* vs. "Operation Description"
White-Space, Formatting, Misspelling: *white-space, formatting, or misspelling*	"Post-operative" vs. "Postoperative"
Word Form Variant: *term is word form variant to preferred or synonym term*	"Preoperative Diagnosis" vs. "Preoperative Diagnoses"
Abbreviation: *term is an abbreviation*	"EBL" vs. "Estimated Blood Loss"
Compound Section Header: *two or more sections designated*	"Operative Indications and Consent"

Same Granularity: *term has same granularity as mapped header*	"Specimens" vs. "Specimens Removed"
Less Granularity: *term has less granularity than mapped header*	"Diagnoses" vs. "Postoperative Diagnosis" or "Preoperative Diagnosis"
More Granularity: *term has more granularity than mapped header*	"Arthroscopic Findings" vs. "Findings"

Each extracted potential section header name was then coded to designate if the header was: the preferred section header name; a new section header from the HL7-ON DSTU; a synonym; had white space, misspelling, or formatting variation; a word form variant; an abbreviation; a compound section header; or a header with additional granularity compared to the HL7-ON DSTU suggested section header specifications. Mapped entries were compared to the section name and coded according to their granularity as: equal, greater, or less granularity. Finally, if a section header could not be mapped to the HL7-ON DSTU, the section header was mapped to LOINC® (21).

Approximately 10% of all mappings were evaluated by both coders (a surgeon and informatician (GM) and a surgeon and informatics graduate student (EA)) in order to assess inter-rater agreement. Percent agreement and Kappa were calculated for mappings to document titles, non-section headers, and section headers; coding for HL7-ON DSTU section headers; and assessment for entry variation (e.g., word forms and synonyms).

The section headers and subsequent mappings were used to create a resource of operative note section headers to improve the reuse of these notes. The resource contains section header terms, term mapping to HL7-ON DSTU and LOINC, and information granularity of mappings.

Results

HL7-ON DSTU Section Header Extraction

Operative note section names and LOINC® section mappings for all designated sections in the HL7-ON DSTU were collected. This included 3 main header elements related to operative note sections which resolved to 6 potential section elements, 12 section names (8 required), and 4 subsection names (all options). Table 2 contains example entries for these 18 elements from the HL7-ON DSTU.

Table 2 –Operative Note Sections from HL7-ON DSTU

Section	LOINC Code	Component Name
Consent (O, H)	N/A	CONSENT FOR SURGICAL PROCEDURE
Anesthesia (R, Sec)	8724-7	SURGICAL OPERATION NOTE ANESTHESIA
Indications (O, Sec)	10217-8	SURGICAL OPERATION NOTE INDICATIONS
Implants (O, Sub)	55122-6	SURGICAL OPERATION NOTE IMPLANTS

R:required;O:optional;H:header;Sec:Section; Sub:Subsection

Operative Note Section Header Evaluation

Automated extraction of headers from 362,311 operative note section resulted in 2,999,414 entries. Removal of entries with a frequency of less than 100 (n=52,054) resulted in 2,947,360 (98.3%) total entries and 476 unique entries.

Initial coding demonstrated that 8 headers (6,975 instances) were document titles, 7 headers (15,525 instances) were document header information, and 77 headers (26,189 instances) did not represent valid potential section headers (Table 3). Of the remaining 384 section headers (2,898,771 instances), 66% section headers (93% instances) mapped to the DSTM and after including LOINC® sections for the remaining elements, successful mappings were obtained for 79% of headers (98% of instances). We also observed large numbers of synonymous terms, normalized variants and other formatting associated with section headers.

Table 3 – Operative Note Section Header Findings

	Headers N (%)	Instances N (%)
Overall	476 (100)	2,947,360 (100)
Document Title	8 (2)	6,875 (0.2)
Header Information	7 (1)	15,525 (0.5)
Not a Section Header	77 (16)	26,189 (0.9)
Section Header	384 (81)	2,898,771 (98)
Map to HL7-ON DSTU	255 (66)	2,735,563 (93)
Granularity		
Same Granularity	179 (70)	2,132,446 (78)
Greater Granularity	65 (25)	594,605 (22)
Less Granularity	11 (4)	8,512 (0.3)
Variation in Terms		
Normalized Word Form	63 (25)	328,090 (12)
Formatting Variation	18 (7)	318,233 (12)
Synonyms	177 (69)	1,203,053 (44)
Abbreviation	18 (7)	256,103 (9)
Map to HL7-ON DSTU or LOINC®	304 (79)	2,833,094 (98)
Mapping Failure	80 (21)	65,677 (2)
Multiple Sections	22 (28)	10,607 (16)
No Mapping	58 (72)	55,070 (84)

Table 4 summarizes mappings to the HL7-ON DSTU including numbers of terms mapping to different headers and the proportion of terms that mapped with equal granularity. There was significant variability in expression for many HL7-ON DSTU section headers, and differences in granularity particularly for section headers for primary performer and secondary performer. An analysis of the 30 most common section terms mapped to the HL7-ON DSTU in all but one case, and the remaining header was a LOINC® section mapping (data not pictured).

Table 4 – HL7-ON DSTU Section Header Mappings

Total Headers	Unique Headers	Section Name	Same Granularity
211,303	4	Anesthesia	99.8%
93,408	6	Complications	100%
16,898	10	Disposition	58.5%
115,307	5	Estimated Blood Loss	100%
2,945	7	Implants	100%
197,127	20	Indications	100%
18,192	20	Operative Note Fluids	79.9%
565,596	33	Operative Note Surgical Procedure	100%
22948	11	Plan	88.8%
283	1	Planned Procedure	100%
327,151	13	Postoperative Diagnosis	100%
333,307	14	Preoperative Diagnosis	99.7%
203,494	21	Primary Performer	0%*
383,174	31	Secondary Performer	0%*
32,530	15	Specimens Removed	99.6%
100,374	28	Surgery Description	100%
1,100	2	Surgical Consent	100%
22,208	2	Surgical Drains	100%
86,415	11	Surgical Operation Note Findings	99.5%
484	2	Surgical Date of Procedure	100%
1319	2	Surgical Procedure Duration	100%

In the overlap coding of 50 entries, percent agreement and Kappa for the initial mapping of document titles, non-section headers, and section headers was 100% and 1.00; the HL7-ON DSTU mapping agreement for section headers mapping was 92% and 0.94, respectively.

A number of section headers did not map to the HL7-ON DSTU or LOINC® section headers. A number of these appeared to be unique to operative notes such as “Tourniquet Time”, “Sponge and Needle Counts”, “Bypass Time”, “Preoperative Antibiotics”, and “Preoperative Status”.

Discussion

As the demand for the extraction of meaningful information from more challenging clinical data sources such as clinical texts becomes an area of greater focus, operative notes and other clinical documents will be reused for a variety of purposes. These efforts aid quality improvement, research, and ultimately clinical data interoperability. This study examines a structured document standard for operative notes, which includes Level 1 (Header) and Level 2 (Section) specifications. Section headers from the HL7-ON DSTU and additional LOINC® sections headers were evaluated on headers extracted from a large number of operative notes from an integrated healthcare delivery system. The standards covered most header instances although amongst unique headers, about 20% did not map. We also observed a large amount of variation in the section header term expression including many synonyms, formatting variations, variation in word forms, and compound section headers within the corpus. Although the large variability of expressions in section headers was not altogether, unexpectedly due to the known variety of expressions commonly seen in clinical text, this was an interesting observation in that the overall section header number was large even for headers, which are generally considered semi-structured and more constrained.

While the HL7-ON DSTU provides eight required section names and a small number of main header items that are conventionally sections in operative notes, our study demonstrates the wide variability in expressions of these elements, the frequent use of optional sections, including the 8 section/subsection elements designated in the HL7-ON DSTU as well as 49 section headers designated in LOINC® and not in the HL7-ON DSTU and 58 section headers unique to both the HL7-ON DSTU and LOINC®. As the HL7-ON DSTU authors note, the base specification for an operative note, like other clinical documents, is the HL7 CDA, Release 2.0, allowing for other sections not present in the HL7-ON DSTU to occur in operative notes. Further, despite the significant challenges with variability in expression of section headers present in our corpus, the “exact text of the section names are not mandated” by the HL7-ON DSTU.

Several of the unique sections that did not map to the HL7-ON DSTU or LOINC®, including “Tourniquet Time”, “Sponge and Needle Counts”, “Bypass Time”, “Preoperative Antibiotics”, and “Preoperative Status” may be valid optional section headers. Some of these are important elements for operative notes for certain subspecialties (e.g., cardiac surgery, transplant surgery, or vascular surgery). Further assessment of operative note sections in other hospital systems may also be helpful for establishing the generalizability of the results of our study.

We also observed issues with respect to both granularity as well as variability in expressing different section headers. In particular, the section header “Disposition” which is standard to the HL7-ON DSTU had a number of entries that were more granular than the general header, including “Postoperative Condition” or “Prognosis”. Similarly, both “Primary Performer” and “Secondary Performer” had wide amounts of variability to the amount of detail expressed. “Primary Performer” included more granular terms such as “Attending Neurosurgeon” and less granular terms such as “Physician”, which is under specified enough that it is unclear whether this represents a Primary Performer or not. Similarly, “Secondary Performer” had mostly more granular terms, many of which were trainees including residents and fellows, as well as the designation of assistants and other providers involved with procedures.

With respect to variability in section header expression, many terms including “Operative Note Surgical Procedure” and “Surgery Description” had many terms to express the same section header. This was similarly the case with “Primary Performer” and “Secondary Performer”, as just described. Surprisingly, sections like “Indications” and “Operative Note Fluids” also had wide variability with 20 different section terms for these two section headers each. We observed that while “Operative Note Fluids” is the section recommended for operative notes, surgeons were sometimes to describe the more significant resuscitative elements like blood products

and colloid administration and instead used ad hoc section headers like “Components Used”.

While the majority of the section headers were fully specified by their name, there were some section headers where the content of the associated section was ambiguous. For instance, the section header “Procedure” or “Procedure(s)” in most cases designates “Operative Note Surgical Procedure”, which lists the procedure(s) performed by the surgeon, similar to the “Surgical Procedure” (Header) which provides coded enumeration of the procedures performed. However, in some cases, the section “Procedure” can be the section most commonly labeled with the operative note section “Surgery Description”, which described the procedure in detail. The disambiguation of these headers may be addressed in future studies at the semantic level with the contents of operative notes using machine-learning or other automated approaches. This also points out the large amount of variability in expression and practice with operative note composition.

There are several limitations to this study. Section headers were extracted from operative notes using a set of deterministic section segmenting rules, and some were likely missing in our analysis. Additionally, while comprehensive in sample size and having data created from academic and community sites, the study was conducted within a single regional hospital system composed of six hospitals including an academic hospital, a children’s hospital, four community hospitals, and three ambulatory surgical centers. The study’s findings could be further validated in a future corroborating study.

Conclusion

Structured document standards and well-formed section header designations are important for interoperability of clinical documents and natural language processing systems that consume these documents. We evaluated the HL7-ON DSTU specification for operative note section headers and LOINC®. Our findings confirm that most section headers are covered by the HL7-ON DSTU and LOINC®. However, there is a large amount of variability in section header expression, and a number of section headers specific to operative notes not currently present in these resources. These findings should be considered for future HL7-ON DSTU iterations and possibly for addition to LOINC®. The resulting section header resource can also be used for section header mappings for natural language processing systems.

Acknowledgments

This research was supported by National Institutes of Health (R01LM011364), Agency for Healthcare Research & Quality (R01HS022085), American Surgical Association Foundation Fellowship, and University of Minnesota Clinical and Translational Science Award (8UL1TR000114). The authors thank Fairview Health Services for support of this research.

References

[1] Singh R, Chauhan R, Anwar S. Improving the quality of general surgical operation notes in accordance with the Royal College of Surgeons guidelines: a prospective completed audit loop study. J Eval Clin Pract. 2011 Jun;18(3):578-80. PubMed PMID: 21210903. Epub 2011/01/08. eng.

[2] Flynn MB, Allen DA. The operative note as billing documentation: a preliminary report. Am Surg. 2004 Jul;70(7):570-4; discussion 4-5. PubMed PMID: 15279177. Epub 2004/07/29. eng.

[3] Lefter LP, Walker SR, Dewhurst F, Turner RW. An audit of operative notes: facts and ways to improve. ANZ J Surg. 2008 Sep;78(9):800-2. PubMed PMID: 18844913. Epub 2008/10/11. eng.

[4] Melton GB. Biomedical and health informatics for surgery. Adv Surg. 2010;44:117-30. PubMed PMID: 20919518. Epub 2010/10/06. eng.

[5] Melton GB, Burkart NE, Frey NG, Chipman JG, Rothenberger DA, Vickers SM. Operative report teaching and synoptic operative reports: a national survey of surgical program directors. Journal of the American College of Surgeons. 2014 Jan;218(1):113-8. PubMed PMID: 24210148.

[6] Mack LA, Dabbs K, Temple WJ. Synoptic operative record for point of care outcomes: a leap forward in knowledge translation. Eur J Surg Oncol. 2010 Sep;36 Suppl 1:S44-9. PubMed PMID: 20609548. Epub 2010/07/09. eng.

[7] Mori AR, Consorti F, Galeazzi E. A tagging system for section headings in a CEN standard on patient record. Proceedings / AMIA Annual Symposium AMIA Symposium. 1998:755-9. PubMed PMID: 9929320. Pubmed Central PMCID: 2232276.

[8] Denny JC, Miller RA, Johnson KB, Spickard A, 3rd. Development and evaluation of a clinical note section header terminology. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2008:156-60. PubMed PMID: 18999303. Pubmed Central PMCID: 2656032.

[9] Meystre S, Haug PJ. Automation of a problem list using natural language processing. BMC medical informatics and decision making. 2005;5:30. PubMed PMID: 16135244. Pubmed Central PMCID: 1208893.

[10] Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform. 2006 Dec;39(6):589-99. PubMed PMID: 16359928.

[11] Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. American journal of human genetics. 2010 Apr 9;86(4):560-72. PubMed PMID: 20362271. Pubmed Central PMCID: 2850440.

[12] Department of Health and Human Services. Centers for Medicare & Medicaid Services. Medicare Learning Network. Evaluation and Management Services Guide. http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/eval_mgmt_serv_guide-ICN006764.pdf. Accessed November 10, 2014.

[13] Joint Commission Operative Note Requirements: Standard IM.6.30, Elements of Performance for IM.6.30. http://www.jointcommission.org/standards_information/jcfaqdetails.aspx?StandardsFAQId=215&StandardsFAQChapterId=43 and http://www.jointcommission.org/NR/rdonlyres/A032623D-02AF-4955-AF7C-08F3D5802E64/0/06_obs_im.pdf. Accessed November 10, 2014.

[14] Centers for Medicare & Medicaid Services (CMS) Operative Note Requirements: Survey Protocol, Regulations and Interpretive Guidelines for Hospitals: A-0396 §482.51. http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/som107ap_a_hospitals.pdf. Accessed November 10, 2014.

[15] Implementation Guide for HL7 Clinical Document Architecture (CDA) Release 2.0, Operative Note Draft Standard for Trial Use (DSTU) Release 1. March 2009. http://www.hl7.org/search/viewSearchResult.cfm?search_id=644266&search_result_url=%2Fdocumentcenter%2Fpublic%2Fstandards%2Fdstu%2F2008SEP%2FCDAR2%5FOPNOTE%5FR1%5FDSTU%5F2009MAR%2Ezip. Accessed November 10, 2014.

[16] Wang Y, Pakhomov S, Melton GB. Predicate argument structure frames for modeling information in operative notes. Studies in health technology and informatics. 2013;192:783-7. PubMed PMID: 23920664.

[17] Wang Y, Pakhomov S, Burkart NE, Ryan JO, Melton GB. A study of actions in operative notes. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:1431-40. PubMed PMID: 23304423. Pubmed Central PMCID: 3540433.

[18] Chen ES, Carter EW, Winden TJ, Sarkar IN, Wang Y, Melton GB. Multi-source development of an integrated model for family health history. Journal of the American Medical Informatics Association : JAMIA. 2014 Oct 21. PubMed PMID: 25336591.

[19] Chen ES, Manaktala S, Sarkar IN, Melton GB. A multi-site content analysis of social history information in clinical notes. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2011;2011:227-36. PubMed PMID: 22195074. Pubmed Central PMCID: 3243209.

[20] Rajamani S, Chen ES, Akre ME, Wang Y, Melton GB. Assessing the adequacy of the HL7/LOINC Document Ontology Role axis. Journal of the American Medical Informatics Association : JAMIA. 2014 Oct 28. PubMed PMID: 25352569.

[21] LOINC 2.48 and RELMA 6.6. https://loinc.org/downloads/files/loinc-and-relma-complete-download/gotoCopyrightedFile. Accessed November 10, 2014.

Address for correspondence

Genevieve B. Melton, M.D., Ph.D. Core Faculty, Institute for Health Informatics, and Associate Professor, Department of Surgery, University of Minnesota, MMC 450, 420 Delaware Street SE, Minneapolis, MN, 55455, USA.

Email: gmelton@umn.edu.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-827

A Process for the Representation of openEHR ADL Archetypes in OWL Ontologies

Alex Mateus Porn[a], Leticia Mara Peres[a], Marcos Didonet Del Fabro[a]

[a] *Informatics department, Federal University of Paraná, Brazil*

Abstract

ADL is a formal language to express archetypes, independent of standards or domain. However, its specification is not precise enough in relation to the specialization and semantic of archetypes, presenting difficulties in implementation and a few available tools. Archetypes may be implemented using other languages such as XML or OWL, increasing integration with Semantic Web tools. Exchanging and transforming data can be better implemented with semantics oriented models, for example using OWL which is a language to define and instantiate Web ontologies defined by W3C. OWL permits defining significant, detailed, precise and consistent distinctions among classes, properties and relations by the user, ensuring the consistency of knowledge than using ADL techniques. This paper presents a process of an openEHR ADL archetypes representation in OWL ontologies. This process consists of ADL archetypes conversion in OWL ontologies and validation of OWL resultant ontologies using the mutation test.

Keywords:

Archetypes, ADL, Ontologies, OWL.

Introduction

An archetype is a formal expression of a distinct concept in the domain level, expressed as constraints on data whose instances are in accordance with the reference model. The development process of an archetype consists mainly of translating a clinical concept for Reference Model entities, defining structure and data representation in an Electronic Health Record (EHR) [28]. Information sources for its development are varied: expert knowledge, specialized publications, data entry screens and system reports, term lists, clinical system data models, messages and regulatory forms, and forms used in clinical appointments or medical records [7, 8].

Clinical archetypes developed according to the openEHR [16] model use the Archetype Definition Language (ADL) as a standard [26]. This language presents an abstract syntax that can be used to express archetypes based on any information model. However, it presents difficulties in implementation and a few available tools to manage contents [3].

Although ADL language has its own syntax, it can be described using other languages such as eXtensible Markup Language (XML) [27] and Web Ontology Language (OWL) [10], increasing the integration with other Semantic Web tools and decreasing interoperability problems between health information systems [3, 4]. Therefore, processes for exchanging and transformating data can be better implemented with semantics oriented models, such as ontologies using OWL [1].

Ontologies define terms to describe and represent a knowledge domain. They are descriptions of concepts and relationships that can be used by people or software agents to share information in this domain. So, they can be used as a unifying structure for semantic representation of information [17, 18].

Ontologies have been frequently used to represent biomedical knowledge in recent years [12, 13]. They provide strict structures for knowledge management tasks related to archetypes and EHR systems representation [3].

We define in this article a process for the representation of ADL archetypes in ontologies described in OWL (or OWL ontologies) which are validated using mutation test. This process has two steps: the conversion of ADL archetypes in OWL ontologies using the algorithm of Elkin et al. [14] and the validation of OWL ontologies using the mutation test of Porn et al. [19, 20].

Archetype Definition Language (ADL)

ADL is a formal language to express archetypes, where the structure is independent of standard or domain. In general, it is not a language for clinical domains. It can be used to define any type of archetype for different reference models therefore it is possible to represent the same syntactic structure among diverse models [1].

ADL archetype is basically composed of three sections [1, 9]: (1) the section *header,* which contains the name of the archetype, a unique code that identifies the clinical concept and information of specialization and the language in which it was written; (2) the *definition* section, which specifies the structure and constraints associated with the clinical concept, restricting the cardinality and the content of the instances of the information model; (3) and the *ontology* section, which represents terminological definition to each clinical concept. In this paper only the d*efinition* section of archetypes is used to the conversion process to OWL.

ADL has some difficulties in implementation. When performing semantic processing two elements are required: one ADL parser and one parser of the reference model to ensure clinical accuracy of content [1].

The ADL parser produces a set of objects without explicit semantic relations among them. The semantics are unknown to the parser and only the association among the elements of *Definition* and *Ontology* section can be defined [1]. Thus, ADL reasoning possibilities are currently very limited [3], as well as the availability of tools to manage ADL contents is reduced. Consequently, the formalization of exchanging data and data transformation processes among systems becomes more difficult than using semantics oriented models, such as OWL ontologies.

Ontology Web Language (OWL)

An alternative to problems related to ADL language is the archetypes representation through OWL ontologies. OWL allows definitions that are meaningful, detailed and accurate with consistent distinctions among classes, properties and

relations defined by the user [1]. Thus, the construction of archetypes using OWL can ensure the consistency of knowledge in a more general way than when using ADL techniques due to the lack of tools for managing these models.

The OWL language is used to define and instantiate Web ontologies and is defined by the W3C (World Wide Web Consortium) [10]. OWL ontology specifies how to derive logical consequences through the formal semantic OWL, clarifying facts that are not presented in the ontology but are bounded by the semantics [11].

Ontologies impose strict structures to knowledge management related to archetypes and EHR systems [3], so that several methodologies and tools have been proposed to compare different ontologies, mixing and identifying inconsistencies. Thus, activities such as comparison, selection, classification and consistency checking can be performed in the OWL in an easier and more efficient way than in the ADL content. As a matter of fact, OWL is a knowledge representation language.

The use of ontologies to represent biomedical knowledge is not a new process, as ontologies have been used in biomedical domains in recent years in different purposes [12, 13]. Thus, OWL becomes appropriate to represent clinical archetypes and information about EHR, so that constraints on archetypes can be determined by OWL or by defining appropriate elements.

Materials and Methods

For the representation of ADL archetypes in OWL ontologies two steps were established:

1. Convert openEHR archetypes developed in ADL language to OWL ontologies, based on an algorithm for mapping ADL objects to OWL, adapted from a model proposed by Elkin et al. [14].
2. Validate OWL ontologies obtained after conversion. For this, OWL ontologies are tested using mutation tests [19, 20].

Figure 1 shows an example of *definition* section of an openEHR archetype implemented in ADL language, that will be converted to an OWL ontology.

```
definition
    EVALUATION[at0000] matches {    -- A health oriented check list
        data matches {
            ITEM_TREE[at0001] matches { -- Tree
                items cardinality matches
                {1..*; unordered} matches {
                    CLUSTER[at0004] occurrences matches
                    {1..*} matches {    -- Question group
                        items cardinality matches
                        {1..*; unordered} matches {
                            CLUSTER[at0002] occurrences matches
                            {1..*} matches {    -- Question
                                items cardinality matches
                                {1..2; unordered} matches {
                                    ELEMENT[at0003] matches {*}
                                    ELEMENT[at0005] occurrences matches
                                    {0..1} matches {    -- A comment on the answer
                                        value matches {
                                            DV_TEXT matches {*}
                                        }}}}}}
                    ELEMENT[at0006] occurrences matches
                    {0..1} matches {    -- Summary
                        value matches {
                            DV_TEXT matches {*}
                        }}}}}
```

Figure 1 – Definition section of an ADL archetype

To represent archetypes in OWL, it is necessary to define in OWL the classes from the Reference Model used by the archetype. Thus, the following algorithm is used to the first step of the conversion process:

1. Reference Model classes used by the archetype are represented as OWL classes;
2. Each archetype node defined as a clinical concept is represented as a subclass of one of the classes of the Reference Model used;
3. Object properties that define associations among nodes and archetype components are defined as OWL Object Properties. The domain declaration of this properties is indicated for the archetype class itself, and the range declaration to the set of objects that should be associated to the domain of this property;
4. Properties that represent data structures in ADL are defined as OWL Data Type Properties, being the domain declaration indicated as the cardinality of objects associated to this property, and the range declaration defined as the data types of all objects found in the archetype with respect to this attribute;
5. Properties constraints and cardinalities defined in ADL archetype are converted in constraints and axioms of OWL language.

Basically, an archetype has a hierarchical structure and constraints started with a root class. Thus, for an ADL archetype that presents several archetype nodes of the Reference Model of the same type, as ELEMENTs, in OWL just one class is defined to represent this node as a root class of the hierarchy and the clinical concepts of the archetype are represented as subclasses of these classes of the Reference Model.

Based on the conversion algorithm, it is possible observe at Figure 2 the definition of the archetype nodes EVALUATION, ITEM_TREE, CLUSTER and ELEMENT of the Figure 1, represented as OWL classes.

```
<Declaration>
    <Class IRI="#CLUSTER"/>
</Declaration>
<Declaration>
    <Class IRI="#ELEMENT"/>
</Declaration>
<Declaration>
    <Class IRI="#EVALUATION"/>
</Declaration>
<Declaration>
    <Class IRI="#ITEM_TREE"/>
</Declaration>
```

Figure 2 – openEHR Reference Model classes in OWL

In the hierarchy of an ADL archetype, properties that define associations among archetype nodes, refer to Information Reference Model classes, therefore they associate individuals of classes from a lower level, to individuals of classes from a higher level, defining classes and subclasses from the archetype [44]. Figures 3 to 7 show excerpts of one OWL ontology representing each structure of ADL archetype of Figure 1. Figure 3 presents OWL Object Properties which represent the properties that define associations among nodes and components of the ADL archetype.

```
<Declaration>
    <ObjectProperty IRI="#data_matches"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
</Declaration>
    .
    .
    .
<Declaration>
    <ObjectProperty IRI="#value_matches"/>
</Declaration>
```

Figure 3 – ADL Operators as OWL Object Properties

In Figure 4 it is possible to observe the domain declarations defined to the OWL Object Properties presented in Figure 3.

```
<ObjectPropertyDomain>
    <ObjectProperty IRI="#data_matches"/>
    <Class IRI="#A_health_oriented_check_list"/>
</ObjectPropertyDomain>
<ObjectPropertyDomain>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
    <Class IRI="#Tree"/>
</ObjectPropertyDomain>
            .
            .
            .
<ObjectPropertyDomain>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#ELEMENT"/>
</ObjectPropertyDomain>
```

Figure 4 – Domain definitions of OWL Object Properties

Figure 5 presents range declarations defined for OWL Object Properties presented in Figure 3.

```
<ObjectPropertyRange>
    <ObjectProperty IRI="#data_matches"/>
    <Class IRI="#ITEM_TREE"/>
</ObjectPropertyRange>
<ObjectPropertyRange>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
    <Class IRI="#CLUSTER"/>
</ObjectPropertyRange>
            .
            .
            .
<ObjectPropertyRange>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#Answer"/>
</ObjectPropertyRange>
<ObjectPropertyRange>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#Summary"/>
</ObjectPropertyRange>
```

Figure 5 – Range Definitions of OWL Object Properties

It is possible to observe in Figure 6 the definition in OWL of the properties that represent data structures in ADL from Figure 1, with domain and range declarations of these properties defined as shown for OWL Object Properties.

```
<Declaration>
    <DataProperty IRI="#value_matches_A_comment_on_the_answer"/>
</Declaration>
<Declaration>
    <DataProperty IRI="#value_matches_Summary"/>
</Declaration>
<Declaration>
    <DataProperty IRI="#value_matches_answer"/>
</Declaration>
```

Figure 6 – ADL Operators as OWL Data Type Properties

As ADL data properties are represented as OWL Data Type Properties according to the conversion algorithm, associations can be made through OWL constructors *hasValue*, *someValuesFrom* and *allValuesFrom*.

So, according to the last step of the conversion algorithm, ADL property constraints and cardinalities are defined as OWL constraints, as shown in Figure 7 the constraints concerning the ELEMENT and ITEM_TREE classes of the archetype from Figure 1.

```
<EquivalentClasses>
    <Class IRI="#ELEMENT"/>
    <ObjectIntersectionOf>
        <ObjectSomeValuesFrom>
            <ObjectProperty IRI="#items_cardinality_matches_question"/>
            <Class IRI="#Question"/>
        </ObjectSomeValuesFrom>
        <ObjectSomeValuesFrom>
            <ObjectProperty IRI="#items_cardinality_matches_tree"/>
            <Class IRI="#Tree"/>
        </ObjectSomeValuesFrom>
    </ObjectIntersectionOf>
</EquivalentClasses>
```

Figure 7 – ADL constraints as OWL constraints

The conversion process and the resulting representation of ADL archetypes to OWL were assisted by a tool for creating and editing of ontologies called Protégé [15]. The following steps were defined:

1. Manual selection of openEHR ADL archetypes in the online repository CKM [16]. Seven archetypes in the review state were selected. Three archetypes are OBSERVATION type, respectively the “Apgar score”, “Autopsy examination” and “Fetal heart rate” archetypes. Two are EVALUATION type, the “Alert” and “A health oriented check list” archetypes. One is INSTRUCTION type and one is an ACTION type, respectively “Informed consent request” and “Follow up action” archetypes.
2. Manual conversion of ADL archetypes to OWL based on the conversion algorithm, being the process realized with the aid of Protégé tool;

To validate archetypes and reveal possible defects inserted in OWL ontologies (proposed in step 2 of the process), we applied the mutation test technique to OWL Ontologies [19, 20].

Results

OWL ontologies obtained after the conversion process are readable for humans and computers, facilitating the interpretation of clinical concepts as well as the development of queries and content management.

Test methods which validate OWL models help to ensure adequacy and quality in ontologies. Tests allow the representation of knowledge domain with security and accuracy. Written queries in SPARQL [25] or DL Query [21] can be used to validate the correction of syntactic and semantic structures in tests of instantiation, recovery, achievement, satisfaction and classification [22].

Published studies have shown that mutation test is the most effective in revealing defects [23, 29]. Mutation test is a defect-based technique proposed in [19, 20, 24] to simulate defects on ontologies, generating mutant ontologies, according to predefined mutation operators known. Mutant ontologies must reveal defects or be considered equivalent (non defective) to the original ontology, according to the DL Query test data used.

After seven openEHR ADL archetypes have been selected in the online repository CKM and converted to OWL, it was possible to generate a total of 2000 mutants ontologies and 187 defects in all OWL ontologies [20].

Detailing in the “Apgar score”, “Autopsy examination” and “Fetal heart rate” archetypes were found respectively 41, 19 and 44 defects. In the “Alert” and “A health oriented check list” archetypes were found 16 and 21 defects and, in the “Informed consent request” and “Follow up action” archetypes revealed 31 and 15 defects. 139 mutant ontologies were defined as equivalent to the original ontology and other mutants were killed with test data generated.

Discussion

Archetypes are used to guide clinical practice, requiring the exploration, comparison, classification and integration of information originating from different heterogeneous systems. The OWL language presents excellent mechanisms for these activities, enabling the representation of ADL archetypes in OWL ontologies and providing an excellent semantic representation of the clinical concepts addressed.

The representation of archetypes in OWL requires the semantic interpretation of clinical archetype, so that the ADL

structures defined in the archetype specialize in openEHR Reference Model classes [9]. Therefore, defects can be created by the developer during the definition of the OWL archetype constraints, and these defects can cause failures. This was possible to observe with the reached results after the application of the mutation test technique in the archetypes was represented in OWL.

It is also possible to observe with the reached results, the large number of mutants generated, determining a high possibility of occurrence of defects in the development. However, the existence of defects in obtained models does not characterize problems in the conversion process. This is because OWL ontologies are based on open world assumption, thus the representation of archetypes in OWL requires the semantic interpretation of clinical archetype.

The purpose of applying the mutation test technique was to validate the obtained results after the conversion process, in this case OWL ontologies, and not the conversion process proposed. Consistently selected archetypes in the online repository CKM were in their review state. Defects were found after the conversion process in OWL ontologies; however, this did not invalidate the proposed method of conversion ADL archetypes to OWL ontologies.

In some cases it was not possible to identify the defect applied by the mutation operator with any given test data because it was not considered in this analysis the instantiation of objects for each archetype classes represented in OWL, which might produce better results. The mutants of these cases are defined as equivalents.

For 139 mutants that have been defined as equivalent, it would be possible to obtain better results (that is, it is possible to define whether the ontology has or not defects) executing the mutant ontologies with same test data and instantiating individuals to ontology classes. Moreover, it can be concluded in these cases that the test data used were not efficient, because they do not produce distinct results from the original ontology and there is the possibility of generating new test data.

With the application of the mutation test on the archetypes in OWL, it is possible to validate the models obtained after the conversion process due to the identification of committed defects, correcting and ensuring the correction of the archetype.

Conclusion

Ontologies define concepts, classes, properties, relationships, constraints and axioms about a particular knowledge domain, which can represent the real and conceptual world through semantic identifiers. Ontologies have proven to be extremely useful to assist the development of computer systems, due to some of their own characteristics such as vocabulary for representing knowledge and the possibility for extending a generic model for a specific domain.

A clinical archetype represents a specific knowledge domain and can be modeled in ontologies in different ways. It was proposed in this paper to convert clinical archetypes of the openEHR standard implemented in ADL to OWL ontologies. In the conversion process, two steps were defined: (1) the conversion from ADL to OWL; (2) the validation of the model obtained.

The ontology developed tool Protégé was used to perform the first step of conversion process of openEHR archetypes to OWL ontologies as well as to perform the second step of process, where it was applied the mutation test in the OWL ontologies obtained. For each mutation performed, it generated a new ontology defined as mutant. For the generation and execution of test data, it used the DL Query language, a query language for ontologies available on the Protégé tool.

For this experiment, seven openEHR archetypes were used, generating an average of 285 mutants to each model. From the total of mutants ontologies obtained, 187 defects were identified and 139 mutants were defined as equivalent to the original ontologies. Each defect revealed in a mutant ontology corresponds to mutation performed by a mutation operator. The analysis and correction were made based on the characteristics of this known mutation operator.

The conversion process was experimented. A test method determined the correction of resulting OWL ontologies. A high number of mutants was generated and defects were revealed, consistent with the fact that they were under review. It is also possible to infer that this test method can be applied to other knowledge domains represented by OWL ontologies.

Acknowledgments

This research was partially funded by Capes.

References

[1] Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT, and Maldonado JA. A model-driven approach for representing clinical archetypes for Semantic Web environments. Journal of Biomedical Informatics, 42(1):150-164, feb. 2009.

[2] Kondo MNS. Mapeamento da base de conhecimento fundamentada em arquétipos: Contribuição à informática em saúde. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 2012.

[3] Martínez-Costa C, Menárguez-Tortosa M, and Fernández-Breis JT. An Approach for the Semantic Interoperability of ISO EN 13606 and OpenEHR Archetypes. Journal of Biomedical Informatics, 43(5):736-746, oct. 2010.

[4] Lezcano L, Sicilia MA, and Rodríguez-Solano C. Integrating reasoning and clinical archetypes using owl ontologies and swrl rules. Journal of Biomedical Informatics, 44(2):343-353, 2011.

[5] Beale T. Archetypes Constraint-based Domain Models for Futureproof Information Systems. OOPSLA 2002 workshop on behavioural semantics, 2002.

[6] Chen R, Garde S, Beale T, Nystrom M, Karlsson D, Klein GO, and Ahlfeldt H. An Archetype-based Testing Framework. Ehealth Beyond the Horizon: Get it there, volume 136 of Studies in health technology and informatics. p. 401-406, Amsterdam, 2008. IOS Press.

[7] Santos MR. Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais. Tese de Doutorado, Universidade Federal de Minas Gerais, Minas Gerais, 2011.

[8] Tortosa MM, and Fernández-Breis JT. OWL-based reasoning methods for validating archetypes. Journal of Biomedical Informatics, 46(2):304-317, 2013.

[9] Kilic O, Bicer V, and Dogac A. Mapping Archetypes to OWL. Software Research and Development Center, 2005.

[10] OWL Web Ontology Language Overview. Available in: http://www.w3.org/TR/owl-features/. Access on: February of 2004.

[11] OWL Web Ontology Language Guide. Available in: http://www.w3.org/TR/owl-guide/. Access on: February of 2004.

[12] Schulz S, and Hahn U. Part-whole representation and reasoning in formal biomedical ontologies. Artificial Intelligence in Medicine, 34:179-200+, 2005.

[13] Smith B. From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. Journal of Biomedical Informatics, 39(3):288-298, jun. 2006.

[14] Elkin PL, Kernberg M, Beale T, Heard S, Shafarman M, Dolin R, and Nelson D. HL7 Template and Archetype Architecture. Template Special Interest Group, nov. 2003. http://www.hl7.org/documentcenter/public/wg/template/HL7_San_Diego_v2.3.zip. Accessed in: November 23rd, 2013.

[15] Protégé Ontology Editor. Available in: http://protege.stanford.edu/. Access on feb 12, 2014.

[16] openEHR. Clinical Knowledge Manager. Available in: http://www.openehr.org/ckm/. Access on: March of 2014.

[17] Davies J, Fensel D, and Harmelen F. Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons, Inc., New York, NY, USA, 2003.

[18] Gruber TR. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Intl J. of Human-Computer Studies - Special issue: the role of formal ontology in the information technology, 43(5-6):907-928, dec. 1995.

[19] Porn AM, and Peres LM. Mutation Test to OWL Ontologies. 13th Intl. Conf. on WWW/Internet, oct. 2014.

[20] Porn AM, Peres LM, and Fabro MD. Teste de Mutação para Arquétipos openEHR Representados por Ontologias OWL. XIV Congresso Brasileiro de Informática em Saúde, Santos, dez. 2014.

[21] DL Query. DL Query Plugin. Available in: http://protegewiki.stanford.edu/wiki/DLQueryTab. Access on: February of 2014.

[22] García-Ramos S, Otero A, and Fernández-López M. OntologyTest: A Tool Evaluate Ontologies through Tests Defined by the User. Springer, p. 91-98, 2009.

[23] Wong WE, Mathur AP, and Maldonado JC. Mutation Versus All-uses: An Empirical Evaluation of Cost, Strength and Efectiveness. Software Quality and Productivity: Theory, Practice and Training, p. 258-265, London, UK, 1995.

[24] Lee S, Bai X, and Chen Y. Automatic Mutation Testing and Simulation on OWL-S Specified Web Services. Simulation Symposium, Annual, 0:149-156, 2008.

[25] SPARQL. SPARQL Query Languagefor RDF. Available in: http://www.w3.org/TR/rdf-sparql-query. Access on: February of 2014.

[26] ADL. Archetype Definition Language. Available in: http://www.openehr.org/downloads/ADLworkbench/learning_about. Access on: February of 2014.

[27] XML. eXtensible Markup Language. Available in: http://www.w3.org/XML/. Access on: February of 2014.

[28] EHR. Electronic Health Record. Available in: http://www.healthit.gov/providers-professionals/faqs/what-electronic-health-record-ehr. Access on: February of 2014.

[29] Budd TA, DeMillo RA, Lipton RJ, and Sayward FG. Theoretical and Empirical Studies on Using Program Mutation to Test the Functional Correctness of Programs. Proceedings of the 7th ACM SIGPLANSIGACT Symposium on Principles of Programming Languages, GIT-ICS, p. 220-233, New York, NY, USA, 1980. Defense Technical Information Center.

Address for correspondence

Informatics department, Federal University of Paraná, Curitiba, Brazil. email: amporn@inf.ufpr.br, lmperes@inf.ufpr.br, didonet@inf.ufpr.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-832

Harmonizing SNOMED CT with BioTopLite: An Exercise in Principled Ontology Alignment

Stefan Schulz[a], Catalina Martínez-Costa[a]

[a]*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract

The integration of heterogeneous ontologies is often hampered by different upper level categories and relations. We report on an on-going effort to align clinical terminology/ontology SNOMED CT with the formal upper-level ontology BioTopLite. This alignment introduces several constraints at the OWL-DL level. The mapping was done manually by analysing formal and textual definitions. Descriptive logic classifiers interactively checked mapping steps, using small modules for increasing performance. We present an effective workflow, using modules of several scales. However, only part of the classes and relations could easily be mapped. The implications for future evolution of SNOMED CT are discussed. It seems generally feasible to use a highly constrained upper-level ontology as an upper level for the benefit of future SNOMED CT versions that are more interoperable with other biomedical ontologies.

Keywords:

Biological Ontologies; SNOMED CT.

Introduction

The support of domain terminologies by formal ontologies is increasingly seen as an important requirement for semantic interoperability in health care and biomedical research. The integration between clinical and research data is essential for the advance of translational and personalized medicine. Both health care and biomedical research have terminology systems. Whereas the Unified Medical Language System (UMLS) [1] has focused on the integration of heterogeneous terminology systems in terms of lexical semantics, the discipline of Applied Ontology [2] has proposed to look beyond linguistic and conceptual structures. Applied Ontology's attention focuses on the referents, i.e., the entities denoted by terms and concepts, their ontological nature and the way they are related. Description logics (DLs) [3], especially the different OWL dialects [4] have become de-facto standards for ontologies.

An important aspect is the standardization of ontology artefacts in terms of upper-level ontologies**,** i.e., formal or semi-formal systems of categories, relations and axioms, linked to human-readable labels. The development of the two most well established upper-level ontologies, namely DOLCE [5] and BFO [6], focused on certain areas like cognitive sciences (DOLCE) and natural science (BFO). Other examples for upper-level ontologies in the biomedical domain include the GALEN upper level [7], the UMLS semantic network [8], and the OBO relation ontology RO [9].

The OBO Foundry effort [10] advocates the use of upper-level ontologies. In parallel, the ontological foundations of the large clinical terminology SNOMED CT [11] have substantially evolved, and the important field of medical classifications, with ICD as flagship, will be increasingly anchored in ontological grounds [12]. Whereas OBO Foundry ontologies are aligned with BFO 1.1 classes and with relations from RO, SNOMED CT's ontological framework is largely influenced by the legacy of former SNOMED versions, together with a frame-like concept model, which evolved almost untouched by formal-ontological deliberations.

An ontological upper-level shared by a range of different domain ontologies would have several advantages:

- Upper-level ontologies would force the categorization of domain entities into well-defined upper level categories. This is not at all trivial, regarding the inherent ambiguity of many legacy categories. A typical example is the pseudo-category *Problem*, a key concept in clinical documentation and decision-making, e.g., as single items that constitute clinical problem lists. *Problem* includes a broad range of statements on a person having a disease now or in the past, having a certain behaviour, having a history of surgical interventions etc., together with epistemic aspects like severity, significance, certainty, priority etc.
- Medical terms that denote what is commonly classified by *Disease* / *Disorder* turn out to be ambiguous [13]. For example, "allergy" may denote a disposition or a process, "fracture" may denote a damaged anatomical entity, but also the fracturing event, or even the life phase during which an organism exhibits the former and which has started with the latter.
- Upper-level ontologies would standardize the ways entities are related with each other. This standardization means a well-defined canon of relations. However, this situation requires a consensus about the precise meaning of relations, in terms of their algebraic properties and relation hierarchies, such as whether the relation **is part of** is reflexive, or whether **is part of** always implies **is included in**. The fact that important relations are ternary (e.g., *a* **is part of** *b* **at time** *t*), whereas the commonly used representation languages are binary is an issue not to be neglected [14]. This fact also complicates the current process of creation of an OWL version out of the upper level ontology BFO2 [15].
- Upper-level ontologies introduce constraining axioms that enforce conformance with domain ontologies that depend on them. Examples include domain / range constraints, such as that only members of the class *Process* can be in the domain of the relation **has participant**, but also further-reaching constraints such as that material entities can have material or immaterial entities as parts, whereas processed can only have processes as parts.

In this paper, we report on a first pilot of a harmonization effort that aims to fill the gap between SNOMED CT and other biomedical ontologies. We use the domain level ontology BioTopLite2 (BTL2) [16] in order to align SNOMED CT with well-defined classes and relations, and we use BTL2's rich set of axioms for validating the design decisions. Finally we will test the scalability of the map by measuring classification time, given that SNOMED CT mapped to BTL2 means to go from polynomial to exponential complexity.

Methods

SNOMED CT is gaining ground as an international clinical terminology standard. After the fusion of earlier SNOMED nomenclatures with the UK clinical terminology CTV3, the international standards organisation IHTSDO has embraced the mission of transforming SNOMED CT into a global healthcare language. Although the ontological SNOMED CT content, to date, is primarily released in relational format, a standard way of creating Description Logics (DL) axioms out of them has been described. This algorithm has been implemented in Perl script, from which the description logics axioms can be generated as an OWL-EL ontology [11]. The DL version has primarily been used to automatically generate taxonomic links in the SNOMED CT production process.

Table 1 shows the concepts of the uppermost level of the SNOMED CT concept hierarchy, under which all of the approximately 300,000 SNOMED CT concepts are grouped.

Table 1 – SNOMED CT upper concepts

Body structure	*Procedure*
Clinical finding	*Qualifier value*
Environment or geogr. location	*Record artefact*
Event	*Situation with explicit context*
Observable entity	*Social context*
Organism	*Special concept*
Pharmaceutical/biologic product	*Specimen*
Physical force	*Staging and scales*
Physical object	*Substance*

Table 2 shows the most frequent relations (object properties) in the OWL version (out of 62). These relations account for approximately 95% of all asserted relational statements in SNOMED CT axioms (totalling approx. 380,000).

Table 2 –most frequent SNOMED CT relations

Role group	**Direct device**
Finding site	**Direct substance**
Method	**Using substance**
Associated morphology	**After**
Procedure device	**Has specimen**
Has active ingredient	**Has focus**
Causative agent	**Finding context**
Has dose form	**Associated finding**
Interprets	**Has intent**
Procedure site - Indirect	**Procedure site**
Direct morphology	**Using access device**
Component	**Procedure context**
Has interpretation	**Associated procedure**
Occurrence	**Access**
Using device	**Specimen source topography**
Temporal context	**Laterality**
Subject relationship context	**Associated with**

BioTop was launched in 2006 as an Upper Domain Ontology in OWL DL. As BioTop never intended to compete with established ontologies, its developers created bridging ontologies to DOLCE, BFO, RO, and the UMLS Semantic Network and left the uppermost level deliberately flat. An important asset of BioTop has been its strong focus on constraining axioms, as an important mechanism for consistency checking, which is not yet available for BFO and RO. BioTop was used as a domain top level in several projects [16]. Later, a reduced form was released, named BioTopLite. Its classes and relations are shown in Tables 3 and 4.

Table 3 – BioTopLite2 upper classes

Disposition	*Process*
Function	*Quality*
Immaterial object	*Role*
Information object	*Temporal region*
Material object	*Value region*

Table 4 – BioTopLite2 upper relations (OWL object properties, without inverses, without descendants)

at some time	**includes**
causes	**precedes**
has condition	**projects onto**
has participant	**represents**

The **mapping process** was done for all SNOMED CT concepts and relations in Table 1 and 2. The use of the complete SNOMED CT OWL version for this exercise would have resulted in severe performance problems, slowing down the whole process.

We therefore decided to perform the mapping on modules of SNOMED CT. One type of modules (M1) was constructed based on signatures that contained one seed concept per SNOMED CT pattern. We defined a SNOMED pattern as a generalization of a subclass or equivalence axioms in the sense that each concept in the axiom was substituted by its uppermost ancestor, i.e., the concepts in Table 1. This method yielded 1,746 axiom types[1]. For each axiom type, a concept was selected randomly. The resulting set of concepts was used as a signature to create a module by following all outgoing links horizontally and vertically, according to [17]. Such a module has about 11,000 classes (variations resulting from random selection have only minor effects). Another series of modules (M2) was created from weighted (by subhierarchies) random signatures of different sizes.

The mapping was done completely manually, following previous work, employing an iterative approach [18,19]. The decision for a given map was done by thoroughly analysing the meaning of the candidate classes and relations, considering formal axioms as well as text definitions and hierarchical context. Each major ontology mapping step is checked by a DL reasoner, the results of which are then analysed and corrected under two perspectives: first, the classes tagged as 'inconsistent' are identified and the causes are investigated and repaired; second, whenever the ontology has reached a consistent state, the logical entailments are analysed for adequacy. Again, the causes are investigated and fixed, whenever wrong

1 Duplications of clauses like '**rel**$_m$ some *TopConcept*$_n$' were ignored; '**rel**$_m$ some *TopConcept*$_n$ and **rel**$_o$ some *TopConcept*$_p$' vs. '**rel**$_o$ some *TopConcept*$_p$ and **rel**$_m$ some *TopConcept*$_n$' and other variations in order were only counted once.

entailments were encountered. We used Protégé 5 [20] for editing, together with the HermiT [21] and FaCT++ reasoners [22], supported by the included explanation facility [23].

Results

The mapping of the classes yielded the following results (BioTopLite classes are indicated by the namespace prefix btl:, SNOMED CT by sct:)[2]:

- **Equivalence mapping** was possible only for the class sct:*Organism*.
- **Simple subclass mappings** were done for the following SNOMED CT top-level concepts: sct:*Event* under btl:*process*, sct:*Observable entity*, *Record artefact* and sct:*Staging and scales* under btl:*information object*, sct:*Pharmaceutical / biologic product*, sct:*Physical object* and sct:*Specimen* under btl:*polymolecular composite entity*, sct:*Physical force* under btl:*quality*, sct:*Procedure* under btl:*action*.
- **Complex subclass mappings**, i.e., those targeting compositional BTL2 expressions were done for sct:*Body structure*, which was mapped to the expression btl:*immaterial object* or btl:*structured biological entity*. Target for sct:*Clinical finding* was the expression btl:*disposition* or btl:*function* or btl:*material object* or btl:*process*. This expression was introduced for convenience as the defined class btl:*condition*, due to the necessity to pragmatically deal with the ambiguous, cross-category meaning of diseases, disorders and findings. sct:*Environment or geographic location* was mapped to btl:*immaterial object* or btl:*poly molecular composite entity*. Finally, sct:*Substance* was mapped to the disjunctive statement btl:*amount of pure substance* or btl:*compound of collective material entities*, due to the fact that also mixtures at molecular or microscopic level (blood, milk) are substances in SNOMED CT.
- **No mappings** resulted from the analysis of (i) sct:*Qualifier value*, (ii) sct:*Situation with explicit context*, (iii) sct:*Social context*, and (iv) sct:*Special concept*. In the first case, we found a large inhomogeneity, including actions like *Training – action*, qualities like *Decreased*, as well as numbers and units of measurement. The second case refers to entities of the frequently debated *Situation with explicit context* hierarchy, which is a kind of SNOMED-internal information model with concepts like *No temperature symptom* or *Treatment changed*. Although most of them could be mapped to btl:*Information object*, we refrained from a mapping. (For a thorough ontological analysis of these so-called context model concepts cf. [24]). Under (iii) we have encountered roles, individual humans, as well as population groups, and under (iv) we found inactive concepts and navigational concepts, which exist only to provide nodes in a navigation hierarchy, according to [25] and are therefore of no ontological relevance.

The mapping of relations proved far more complex, because of their larger number in both ontologies (see Tables 2 and 4) and the large number of constraining axioms attached to BTL2 relations. We decided to map not only the relations at the uppermost hierarchical level but also some of their descendants, because of their frequent use and special semantics. So far, we have limited the relation mapping effort to those relations that, together, cover 95% of all relational statements in SNOMED CT, see Table 2.

Only one relation equivalence was found, namely sct:**After** for btl:**precedes**.[3]

A special case was the sct:**RoleGroup** relation, which had originated as a mere syntactic construct to circumvent nested expressions [26], and then was found to be interpretable in several ways according to its context. We identified two different possible mappings: for domains of the type sct:*Clinical Finding* the corresponding BTL relation is btl:**has condition**, domains of the type sct:*Procedure* would be mapped to btl:**has part**.

Most SNOMED CT relations could be mapped as subrelations to BTL2 relations, together with refined domain and range restrictions. This is the case with the following relations:

- sct:**Finding site** mapped to btl:**is included in** with domain btl:*condition* and range sct:*Body structure;*
- sct:**Procedure site** mapped to btl:**is included in** with domain sct:*Procedure* and range sct:*Body structure;*
- sct:**Using device** mapped to btl:**has patient** with domain sct:*Procedure* and range sct:*Physical object*;
- sct:**Causative agent** mapped to btl:**caused by** with range sct:*Substance*;
- sct:**Laterality** mapped to btl:**bearer of** with range btl:*Quality*;
- sct:**Has active ingredient** mapped to btl:**has part** with domain sct:*Pharmaceutical / biologic product* and range sct:*Substance*;
- sct:**Associated morphology** mapped to btl:**has part** with domain btl:*condition* and range sct:*Morphologically abnormal structure*;
- sct:**Has dose form** mapped to btl:**is bearer of** with domain sct:*Pharmaceutical / biologic product* and range sct: *Drug dose form*;
- sct:**Has specimen** mapped to btl:**has patient** with domain sct:*Procedure* and range sct: *specimen*;
- sct:**Method** mapped to btl:**includes** with domain sct:*Procedure* and range sct:*Qualifier value*;
- sct:**Access** mapped to btl:**is bearer of** with domain range sct:*Qualifier value*;
- sct:**Occurrence** mapped to btl:**projects onto** with domain range sct:*Qualifier value*;
- sct:**Procedure device** mapped to btl:**has patient** with domain sct:*Procedure* and range sct:*Physical object*;
- sct:**Component** mapped to btl:**has patient** with domain sct:*Procedure* and range sct:*Substance*;
- sct:**Specimen source topography** mapped to the chain (btl:**at some time** • btl:**is included in**).

[2] We use *Italics* for classes / concepts and **Bold face** for relations (object properties)

[3] Another relation pair for which equivalence mapping seems plausible is sct:**part of** / btl:**is part of**. The former is not in our list as it is, to time, is infrequent SNOMED CT**.** After the planned redesign of the anatomy branch in SNOMED CT, this will be one of the most frequent relations.

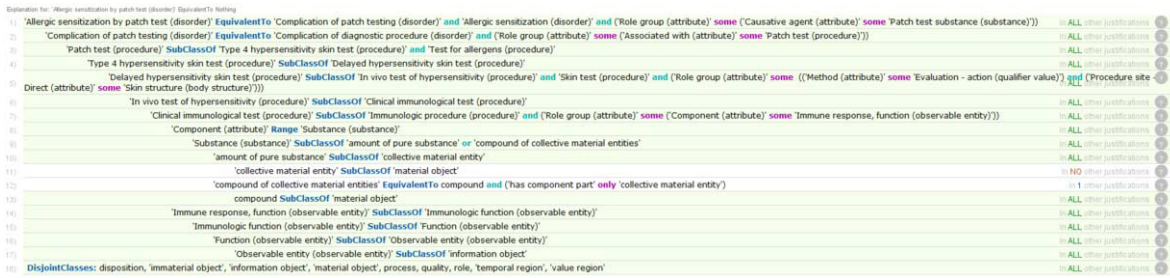


Figure 1 – Documentation of an unsatisfiable class using the Protégé explanation function

Several SNOMED CT subrelation introduce additional aspects, e.g. sct:**Procedure site - Indirect**, sct:**Direct morphology**, sct:**Direct device**, sct:**Using access device**. The differentiation of their meaning regarding their respective superrelations cannot be expressed by any BTL relation. One solution would be to assign roles or qualities to the target classes, e.g., the difference between direct or indirect surgical access.

There are complex relationships in SNOMED CT that cannot be mapped to any BTL2 relation. For instance sct:**Has focus** describes the intent of a procedure to reach a certain goal. An appropriate representation would require a model of intentionality, together with a reference to a class that is not (yet) instantiated. This would require an approach similar to the representation of plans, which is outside of what can be expressed in OWL-EL [27].

As argued in [24], the concepts in the SNOMED CT hierarchy *Situation in specific context* represent epistemic entities like components in an information model. As shown in [28], there are possibilities to express such statements in description logics, as instances of the class btl:*information object*, however requiring a more expressive logic than OWL-EL. We therefore exclude the following relations from our mapping: sct:**Finding context**, sct:**Procedure context**, sct:**Associated finding**, sct:**Interprets**, sct:**Has intent**, sct:**Subject relationship context**, sct:**Has interpretation**, sct:**Temporal context**, sct:**Subject relationship context**.

The mapping workflow turned out to critically depend on classification time. Whenever the ontology was in an inconsistent state, the classification of SNOMED CT modules of type M1 (in which all SNOMED CT patterns are represented) causes disruptions of more than fifteen minutes. In such cases, a switch to M2 modules of smaller scale was mostly sufficient for debugging, with classification times under two minutes. The debugging of insatisfiable classes was time consuming, which is illustrated by the output of the Protégé explanation function (Figure 1). At the end of the described phase, with a mapping of 95% of the relational clauses, there were on average three insatisfiable SNOMED CT concepts in modules with on average 11,050 concepts.

Conclusion

The on-going SNOMED CT – BTL2 aligning experience has shown up until now that large parts of SNOMED CT were compatible with a highly principled and compact upper level ontology like BioTopLite. The mapping at class level demonstrated a good agreement with BioTopLite for fifteen of the nineteen subhierarchy roots. That ontologically heterogeneous subhierarchies like sct:*Situation with specific context*, sct:*Qualifier value*, as well as sct:*Social context* and sct:*Special concept* could not be straightforwardly mapped was not surprising. We recommend that these SNOMED CT branches undergo a major redesign. At the level of the relations, the known ambiguity of the role group relation was confirmed. This construct requires further analysis and should eventually be substituted by a set of new relations that are clearly labelled. However, this would have to include nested relations into the SNOMED CT architecture, a desideratum that had been formulated on various occasions, by several stakeholders, such as the IHTSDO Observable group.

Regarding the other relations, SNOMED CT would be well served in most cases with a much smaller set of relations. The multitude of relations is, above all, a legacy issue. The simple BTL2 relation btl2:**is included in** would perfectly suit to substitute the SNOMED CT relations sct:**Finding Site** and sct:**Procedure Site**. The same would be true for btl:**has part**, which could substitute the relations sct:**Active ingredient** and sct:**Associated Morphology**. From an ontological point of view, we see no need to repeat the sort of the range concept (e.g., being a morphology or a substance) in the relation. This would simplify the construction and maintenance of SNOMED CT.

Although BTL2 uses the whole range of OWL-DL constructors, classification of SNOMED CT modules under BTL2 even of a size of 15,000 classes show satisfactory performance values, with about 15 minutes using the HermIT classifier. For a non-disruptive workflow it has been proven valuable to use small random modules that classify quickly after each modification of the map, because it is quite likely that the impact of a modelling error leads to insatisfiable classes even in these small modules. Once all small modules are satisfiable, then the more time consuming consistency check with a big module can be performed. Such modules that cover the complete variability of SNOMED CT by containing at least one class for each occurring design pattern have a size of approximately 11,000 classes. Checking the map with these modules might spot additional errors, or confirm the satisfiability documented by the smaller (incomplete) modules.

This approach could be integrated into the SNOMED CT maintenance and redesign workflow, which until now is mainly guided by the constraints formulated inside the frame-like SNOMED CT concept model. In contrast, the OWL version used in the production process cannot spot inconsistencies due to the inexpressiveness of the language.

In 2008, Rector and Brandt argued in favour of a more expressive description logics for SNOMED CT [29]. Our findings suggest the feasibility for this at least regarding the upper level.

Acknowledgements

This work has been funded by the SemanticHealthNet Network of Excellence within the EU 7th Framework Program,

Call: FP7-ICT- 2011-7, agreement 288408. http://www.semantichealthnet.eu.

References

[1] United States National Library of Medicine. Unified Medical Language System (UMLS), 2015. http://www.nlm.nih.gov/research/umls.

[2] Munn K, Smith B, eds. Applied Ontology. An Introduction, Frankfurt am Main: Ontos, 2008.

[3] Baader F et al. The Description Logic Handbook (2nd Edition). Cambridge: Cambridge University Press, 2007.

[4] OWL2 profiles http://www.w3.org/TR/owl2-profiles.

[5] Borgo S, Masolo C. Ontological Foundations of DOLCE. Handbook on Ontologies (2nd Edition), Springer Verlag, 2009: 361-382.

[6] Grenon P, Smith B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation 4 (1): 69-103.

[7] Rector A, Rogers J. Patterns, Properties and Minimizing Commitment: Reconstruction of the GALEN Upper Ontology in OWL. Proc. of the EKAW*04 Workshop on Core Ontologies in Ontology Engineering, 2004, http://ceur-ws.org/Vol-118/.

[8] McCray AT. An upper-level ontology for the biomedical domain. Comp Funct Genomics. 2003; 4 (1): 80-84.

[9] Smith B et al. Relations in biomedical ontologies. Genome Biology 2005; 6 (5): R46. Epub 2005 Apr 28.

[10] Smith B et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 2007 Nov; 25 (11): 1251-1255.

[11] International Health Terminology Standards Development Organisation. SNOMED CT® Technical Implementation Guide. Copenhagen, Denmark, 2015. http://www.snomed.org/tig.pdf.

[12] Rodrigues JM et al. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. Stud Health Technol Inform. 2013;192:343-346.

[13] Schulz S et al. Scalable representations of diseases in biomedical ontologies. J Biomed Semantics, 2011, 17;2 Suppl 2:S6.

[14] Zamborlini V, Guizzardi G: On the Representation of Temporally Changing Information in OWL. EDOCW 2010: 283-292.

[15] Smith B et al. Basic Formal Ontology 2.0. Draft specification and user's guide. http://purl.obolibrary.org/obo/bfo/2012-07-20/Reference .

[16] Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In: Furbach U, Staab S eds. Informatik 2013. IOS Press; 2013.

[17] Seidenberg J, Rector AL: Web ontology segmentation: analysis, classification and use. WWW 2006: 13-22.

[18] Schulz S, Hahn U. Medical knowledge reengineering – converting major portions of the UMLS into a terminological knowledge base. Int J Med Inform. 2001;64:207-221.

[19] Schulz S et al. Alignment of the UMLS semantic network with BioTop: methodology and assessment. Bioinformatics. 2009 Jun 15; 25 (12): i69-76.

[20] Stanford Center for Biomedical Informatics Research. The Protégé Ontology Editor and Knowledge Acquisition System, 2015. http://protege.stanford.edu.

[21] Shearer R, Motik B, Horrocks I, HermiT: a Highly-Efficient OWL Reason-er. 5th OWL-ED Workshop, 2008.

[22] Tsarkov D, Horrocks I. FaCT++ Description Logic Reasoner: System Description. In Third International Joint Conference on Automated Reasoning, 292-297. Springer, 2006.

[23] Horridge M, Parsia B, Sattler U. Laconic and precise justifications in OWL. Proc. of ISWC-08, volume 5318 of LNCS, 323-338, 2008.

[24] Martínez-Costa C, Schulz S. Ontology-based reinterpretation of the SNOMED CT context model. Proc. of the 4th International Conference on Biomedical Ontology, CEUR Workshop Proceedings, 2013; 1040: 90-95.

[26] Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED. Proc AMIA Symp. 2002:712-716.

[27] Schulz S, Cornet R, Spackman K. Consolidating SNOMED CT's ontological commitment. Applied Ontology, 2011; 6: 1-11.

[28] Schulz S et al. An ontological analysis of reference in health record statements. FOIS 2014, 289-302.

[29] Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. J Am Med Inform Assoc. 2008 Nov-Dec;15(6):744-751

Address for correspondence

Stefan Schulz
Institut für Medizinische Informatik,
Statistik und Dokumentation
Medizinische Universität Graz
Auenbruggerplatz 2/V
8020 Graz
Austria
stefan.schulz@medunigraz.at

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-837

Standardized Cardiovascular Quality Assurance Forms with Multilingual Support, UMLS Coding and Medical Concept Analyses

Julian Varghese[a], Sarah Schulze Sünninghausen[a], Martin Dugas[a]

[a]*Institute of Medical Informatics, University of Muenster, Germany*

Abstract

Standardized quality assurance (QA) plays an import role to maintain and develop success of cardiovascular procedures (CP). Well-established QA models from Germany could be shared in a form repository for world-wide reuse and exchange.

Therefore, we collected the complete set of all quality QA forms for CP, which is obligatory to be filled out by all German health service providers. Original forms were converted into standardized study forms according to ODM (Operational Data Model) and translated into English. Common medical concepts and clusters of medical concepts were identified based on UMLS coding of form items.

All forms are available on the web as multilingual ODM documents. UMLS concept coverage analysis indicates 88% coverage with few but critically important definition gaps, which need to be addressed by UMLS.

Keywords:

Cardiovascular Procedures; Quality Assurance; Medical Forms; UMLS.

Introduction

Cardiovascular diseases (CD) represent one of the leading causes of death in adults in the majority of developed countries [1] and are already rising in poorer populations [2]. Patients suffering from severe cases of CD, which cannot be treated solely by medication or conservative therapy, need to receive surgical or non-surgical interventions as lifesaving treatments.

Existing methods of quality assurance (QA) are necessary to measure, maintain, and further develop the medical quality of medical procedures among different health service providers. Since 2000, all German health service providers are obliged by law to apply QA for their services [3]. A distinct form-based QA documentation has evolved for 30 medical specialties containing more than four million patient cases, indicating outstanding efforts in German QA at international level [4]. This QA documentation also poses major challenges for routine hospital documentation, which already represents a significant amount of physicians' work time. The original QA-forms, which are already in use, are defined in a structured way but do not apply standardized data models.

Innovative approaches to interoperable medical forms need to be established to incorporate those forms into different hospital information systems for harmonious structured clinical data storage and linkage to routine documentation. To enforce this, it is important to use existing standard data models and semantic codes from international terminologies for data items within those forms. Thus, each data item is defined in a structured way and linked to a medical context unambiguously such that form items get machine-readable and language-independent. The first feature is beneficial for clinical data management and data analysis where electronic forms are used for medical documentation and form-based data retrieval [5]. The last feature enables the world-wide reuse of medical forms where QA is less developed.

To achieve this, international vocabularies with a large coverage for clinically relevant medical terms need to be used such that medical terms (e.g. Myocardial infarction, Dissection of aorta) can be correctly mapped semantically to data elements in the forms.

The objectives of this work are threefold:

1. Acquire all current German QA forms for cardiovascular procedures (CP) and process them into the standardized model of study forms: Operational Data Model (ODM) [6] specified by the Clinical Data Interchange Standards Consortium (CDISC) [7] and supported [5] by regulatory agencies such as Food and Drug Administration (FDA) [8] and European Medicines Agency (EMA) [9].
2. Translate German data items into English and semantic coding using the Unified Medical Language System (UMLS) [10] based on medical expert decision. All forms will be accessible on a medical form repository, which has been developed, based on our previous work [11].

   This way, we will provide the first medical E-forms for QA, with structured data items and international semantic annotations in a technical standard format and open access to enable international discussion, editing and re-use.

   Based on the English translation and the language-independent medical context of items provided by UMLS codes, further languages can be easily added.

   To give an example of our methods of form sharing and multilingual support, a clinical Follow Up form has been published previously with semantic coding and English translation, upon which currently 17 languages (German, English, French, Portuguese, Spanish, Mandarin, Hindi, Japanese, etc.) have been added [12].
3. Based on UMLS-coding of form items, frequency analysis of medical concepts is conducted to identify most frequently used concepts and clusters of concepts within the domain of cardiovascular diseases relevant to QA. Furthermore, a coverage analysis of UMLS will show the possible extent of critical concept definition gaps for medical QA.

Supplement files including full data tables and large figure images are available on:

https://drive.google.com/folderview?id=0BxfnhHTIk8tqSlJLb3BZUUpGamM&usp=sharing

Methods

Forms retrieval

All available QA forms for CP have been acquired from the German Institute for Applied Quality Improvement and Research in Health Care (AQUA) [13]. Permission for scientific analysis and publication was granted.

The original forms were provided as PDF files and a data model for all forms using a relational Microsoft Access database [14]. Thus, structured information of every data item (as item name, datatype, measuring unit, min/max value, mapping to code lists, whether or not an item is mandatory) is available.

Thirteen forms were retrieved, which represent the complete set of CP which are obliged to be filled out by every German health care facility that provides CP. Table 1 lists procedure names and their associated diseases.

Table 1- PCI = Percutaneous coronary intervention, Tx = Transplant, ICD = Implantable Cardioverter Defibrillator

Procedure	Treated Main Diseases
Coronary angiography, PCI Coronary Surgery	Heart failure, Heart attack/Myocardial infarction, Coronary Heart Disease, Arterial Hypertension
Pacemaker Implantation Pacemaker Replacement Pacemaker Revision/Expl. ICD Implantation ICD Replacement ICD Revision/Expl.	Cardiac arrhythmias, Diseases affecting Heart's electrical conduction system
Carotid Revascularization	Cerebrovascular accidents / Strokes, Arteriosclerotic Diseases
Heart Transplant, Cardiac assistance Heart Transplant Follow-Up Heart+Lung Tx Heart+Lung Tx Follow-Up	Heart failure, Pulmonary hypertension, Cystic fibrosis, Congenital heart and lung affections

Form conversion and semantic enrichment

A converter was implemented to convert the original forms into ODM study forms according to the current ODM specification 1.3.2. Some of the original forms consist of sub-forms (e.g., QA form for Heart Transplant and Cardiac assistance consists of a "Follow Up"-, "Basic documentation"- and a "Procedure details" part). For conversion, a single sub-form is represented by one ODM-Form. Sub-forms belonging to one form share the same title prefix.

Altogether, the 13 original forms represent 24 forms including sub-forms.

A manual review of all original forms was carried out by three medical experts, including one physician with certified English proficiency (IELTS Band 7.5) to first translate German into English item names and then apply semantic coding using the UMLS Metathesaurus (version 2014AB) based on consensus decision.

Due to variations in concept granularity and some existing concept duplicates within UMLS that cause ambiguities when mapping medical terms into UMLS codes, we make use of previously published coding principles for pre-coordinated and post-coordinated concept codes [15]. Thus, form data items are assigned to UMLS codes with medical proficiency and previously used codes will be reused wherever possible. Data items whose values can be directly inferred from previous data items within the same form will not be coded. For example, the items "Year of Implantation" and "Year of Implantation is unknown" will be coded only once.

It has to be noted, that only the concept domain was regarded when mapping data items to medical concepts but not the value domain. Therefore, existing list items of a data item will not be taken into account for translation and semantic coding.

After having assigned UMLS codes, the medical context of each data item is defined language-independently. For many medical concepts explicit language translations are already available within UMLS [16]. The explicit translated language expression of a data item is defined by a "TranslatedText" element within the item definition according to ODM standard.

UMLS coverage analysis

During the process of semantic coding, all medical concepts that cannot be mapped with sufficient clinical specificity (according to coding principles, [15]) will be collected. These concepts will be marked with internal codes for identification.

Frequency Analysis of medical concepts

To provide a list of common data items, overall frequency of medical concepts that represent all data items will be calculated by counting same UMLS codes. Cumulative frequency analysis will illustrate cumulative coverage of most common concepts among the complete set of all concept occurrences within the acquired QA forms.

Clinical categories of most common concepts will be manually identified to represent the most common concepts within a hierarchical heat map tree based on manual expert review.

Thus, clinical categories are represented by tree nodes, which are based on hierarchical relations based on previous work [15] to cluster most common concepts in clinical trials.

Form sharing

All QA forms with semantic UMLS annotations will be uploaded to the forms repository [11] with open access for all registered users (registration is free).

The repository will provide form visualization and download, user-ratings, user comments, form-edition and form-versioning.

Results

Form conversion and UMLS coverage

Based on 13 original QA forms, 24 ODM forms have been generated with semantic codes for every data item wherever UMLS codes are available. The resulting 24 forms contain

1011 data items referring to 796 medical concepts of which 361 are unique.

Among those 796 medical concepts, 702 could be coded via UMLS codes (Concepts with missing codes: 94, Coverage 88%). The 94 medical concepts with missing UMLS-codes were semantically annotated with internal identifiers to count them for the frequency analysis of all concepts. Table 2 represents an extract of 5 concepts we deemed critically important, which are not defined by UMLS.

The full list of concepts is available in the supplement file: Coverage.xlsx.

Table 2- Critically important medical concepts that occur in QA forms, which are not covered by UMLS. PCI = Percutaneous coronary intervention

Concept Name	QA form(s)
1. Applied Dose Area Product (X-Ray exposure)	PCI, Coronary Surgery, ICD/Pacemaker Procedures, Heart Transplant, Cardiac assistance
2. Quality of transplanted organ at time of harvesting / transplantation	Heart/Heart-Lung Transplant
3. Door-to-Balloon Time in myocardial infarction	PCI
4. Lung Allocation Score	Heart-Lung Transplant
5. Multilevel lesions of carotid artery	Carotid Revascularization

Frequency Analysis of coded medical concepts

Tables 3 and 4 provide a top five and top ten extract of the most frequent concepts for demographic/administrative and clinical concepts, respectively. Both tables present for each concept its respective UMLS code (CUI) and relative concept frequency (RCF), which is the ratio of the absolute concept frequency of that concept to the number of all concepts occurring in the QA forms. For example, the concept "Date of Birth" occurred 15 times, since the total number of concept occurrences in all forms is 796, RCF=1,89%.

The clinical concept "Procedure code" has the highest RCF (3.89%), since all forms focus on cardiovascular procedures and some forms require more than once data items, which are conceptually related to the procedure code.

For the full list of concepts and their UMLS codes and absolute frequencies, see supplement file: FrequencyAnalysis.xlsx.

Table 3– Most frequent demographic/administrative concepts CUI: Concept unique identifier, RCF= Relative concept frequency

Concept Name	CUI	RCF (%)
1. Date of Birth	C0421451	1,89
2. Medical specialty of health care facility	C0037778	1,89
3. Gender	C0079399	1,63
4. Facility section identifier	C1547540	1,63
5. Zip code of discharge	C1521842	1,63

Table 4– Most frequent clinical concepts

Concept Name	CUI	RCF (%)
1. Procedure code	C1550373	3,89
2. R wave height	C0429049	1,89
3. (Discharge/Admission) Diagnosis	C0011900	1,76
5. Manufacturer Name (of medical device)	C0947322	1,50
6. Reason for Patient Discharge	C3164834	1,38
7. Number of times, planned procedure has been applied before (procedural record)	C2114712	1,13
8. ASA classification	C1531480	1,13
9. Procedure duration	C1442476	1,00
10. Applied Dose Area Product (X-Ray)	-Not available-	0.88

Figure 1 illustrates the cumulative coverage of all 361 unique concepts starting with the most frequent concept "Procedure code" (rank 1, frequency = 31) on the left hand side of the graph, followed by the remaining frequency-ordered concepts.

Blue circles indicate slope changes of the graph due to frequency reduction. The red circle represents the 112th most frequent concept, which has an absolute frequency = 2. All other concepts on the remaining right side of the graph share an absolute frequency = 1. Thus, it can be seen that by only using a third of all unique concepts (112/361), nearly 70% of all concepts occurrences in the QA-forms can be covered.

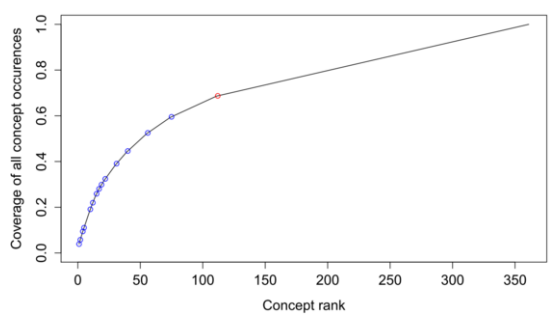


Figure 1– Cumulative frequencies, starting left with most frequent concept 'Procedure code' (rank =1, frequency= 31), blue circles indicate slope reductions, red circle marks concept rank 112, after which concept frequency = 1

Figure 2 (larger image available in the supplement) shows the category tree depicting clinical categories and subcategories fitting these 112 most frequent medical concepts. The color of a node indicates the number of concept occurrences within the QA forms that fall into the category represented by that node, similar to a heat map. Every category is annotated with a count number in parenthesis, which states the exact number of times medical concepts from the QA-forms fall into that category. Note that the count number of a parent's node doesn't need to be the sum of its child nodes' count numbers, since the parent node can contain concepts, which cannot be assigned to any of its child nodes.

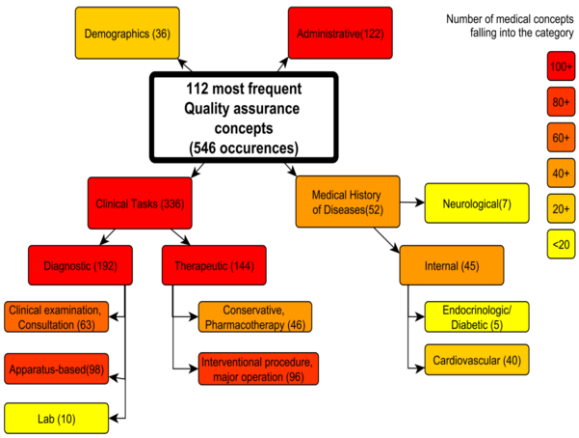


Figure 2– Categorical Heat Map Tree categorizing all 546 concept occurrences for concepts that have at least a frequency = 2. Refer to the text section for detailed description.

Medical categories as Diagnostic/Therapeutic processes and administrative processes encompass the most frequent medical concept occurrences for documentation of QA forms. Compared to our previous work [15] to cluster most common concepts for clinical trials the main category "Administrative" was used instead of "Trial/Research-specific concepts" to represent administrative patient-related QA data items like "Health care facility identifier" or "Zip code of discharge".

"Neurological diseases" is a further category next to "Internal medicine diseases", which contains some of the medical concepts (n=7) in the QA forms, since neurovascular accidents play an import role with people receiving cardiovascular procedures.

For the full list showing which concept was mapped to which category, refer to the supplement file: FrequencyAnalysis.xlsx.

Form sharing

All ODM forms are accessible on the form portal [11]. A search function for retrieving the forms presented in table 1 is provided. The exact names of form titles are given in the supplement file: Form-Names.xlsx. Forms can be downloaded as native ODM files and/or visualized on the website.

Figure 3 shows an extract of a detailed view on the item group "Operation" from the QA Form "Pacemaker replacement" containing the data items "Operation date", "Procedure codes" and "Procedure Duration" on the form portal.

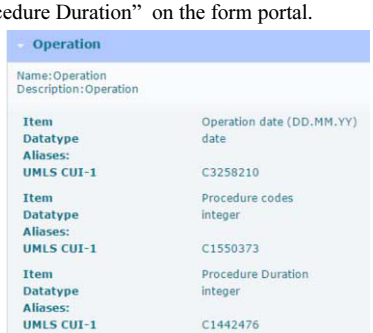


Figure 3– Detailed view of item group "Operation" on the form repository for the QA form "Pacemaker replacement" showing item names and semantic annotations.

Discussion

Conversion and Translation of Forms

Conversions between different formats are associated with a certain degree of loss of information. Since format conversion was applied from one structured format into another standardized structured (ODM) format this process could be implemented automatically. ODM at the current specification 1.3.2 provides the full amount of item structuring and item definition features to carry out the conversion without any losses of content relevant to medical documentation (e.g., layout-specific information or internal abbreviations as additional information was ignored).

UMLS coding and UMLS coverage

The analyzed QA forms are based on obligatory QA implementation in German hospitals since the year 2000 and established by AQUA [13] (commissioned by the German Federal Joint Committee) as a long-term consensus-driven result by medical specialists (doctors, nurses, health scientists, pharmacists, economists, statisticians, IT specialists etc.) [3, 4, 13]. Hence, data items of those forms have an evidential basis for medical significance in routine documentation to assess, maintain and improve medical quality for medical treatment.

Therefore, our results regarding medical concepts that represent data items of the QA forms and are not defined by UMLS indicate significant definition gaps. Even though most of the concepts occurrences could be covered, 12% remain undefined and are critically important to be incorporated into UMLS or other large medical vocabularies to make those concepts amenable for unambiguous identification, which is an important prerequisite for data harmonization and interoperability of medical data items.

Medical Concepts and Form Data Items

Calculated frequencies and established clustering of concepts are based on the mapping of all data items from the QA forms into medical concepts. It has to be noted that medical concepts represent the concept domain of a data item and not necessarily further information as value-domain-specific information or temporal information.

If the user is interested in which items are linked to concept codes our published form files can be searched for a given UMLS code, which is defined within a data item definition element. Furthermore a data item definition element contains information on measurement units, min and max values wherever this information was given in the original forms.

Limitations

Expert-based manual coding of data items is a time-consuming process, application of our methods on a larger set of form data items leads to an increase of coding efforts linear to the size of the item set. Carrying out multiple manual coding by different coders is possible, however it has to be noted, that UMLS is not a classification, it is a meta-vocabulary with some concept duplicates and varying concept granularities leading to potentially different code mappings among different coders. Since we apply our methods on previously published coding principles [15] and a very limited set of different medical experts as coders with consensus-based code mappings, identification of different medical concepts is based on medical expert knowledge and the mentioned UMLS coding issues could be counteracted.

Translation from German into English items, which is the prerequisite for mapping into language-independent UMLS codes, was limited to only one medical expert as a translator

with certified IELTS-English Proficiency (Band 7.5). A certified medical translator was not involved in this work. However, since the work is based on the translation of data items with a limited sequence of words and translation is carried out by a medical expert, content and context of medical translation should not have produced misleading results. Additionally, a platform is provided for international discussion and reuse enabling registered user to suggest editions of the currently available forms. With the provided web-portal infrastructure and an integrated web-based ODM form editor, changes of translation and new languages additions (including code list items) are supported without losing the content structure. For illustration purposes, see [12].

Conclusion

The work provides the first E-forms for quality assurance for cardiovascular procedures containing language-independent data items with semantic codes. A platform is provided for open-access to enable world-wide reuse and versioning.

Frequency analysis and clusters of medical concepts that represent data items provide a set of conceptual common data elements. It could be shown, that by only using around 30% of all unique medical concepts, nearly 70% of all concept occurrences could be covered when only focusing on the most frequent concepts. Therefore, these most frequent concepts should be given special consideration when forms relevant to cardiovascular QA documentation are being established.

UMLS-coverage analysis indicates concept coverage of nearly 90% with few but critically important concept definition gaps. A UMLS update of the presented concepts is therefore strongly recommended.

References

[1] Sans S, Kesteloot H, Kromhout D. The burden of cardiovascular diseases mortality in Europe. Task Force of the European Society of Cardiology on Cardiovascular Mortality and Morbidity Statistics in Europe. European Heart Journal, 1997: 18; pp. 1231-1248.

[2] Nguyen QN, Pham ST, Do LD, Nguyen VL, Wall S, Weinehall L, Bonita R, Byass P. Cardiovascular Disease Risk Factor Patterns and Their Implications for Intervention Strategies in Vietnam. International Journal of Hypertension, 2012: Article ID 560397.

[3] Bramesfeld A, Willms G. Cross-Sectoral Quality Assurance, § 137a Social Code Book V. Public Health Forum, 2014: 22(2); pp. 14.e1-14.e3

[4] Heller G, Szecsenyi J, Willms G, Broge B. Quality measurement using administrative data in mandatory quality assurance. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2014: 108(8-9); pp. 465-469.

[5] Dugas M, Fritz F, Krumm R, Breil B. Automated UMLS-based comparison of medical forms. PloS One, 2013 July 4: 8(7).

[6] Operational Data Model. Available from: http://www.cdisc.org/odm Accessed 2014 Dec 12.

[7] Clinical Data Interchange Standards Consortium. Available from: http://www.cdisc.org/odm Accessed 2014 Dec 12.

[8] U.S. Food and Drug Administration. Available from: http://www.fda.gov Accessed 2014 Dec 12.

[9] European Medicines Agency. Available from: http://www.ema.europa.eu/ema . Accessed 2014 Dec 12.

[10] Unified Medical Language System®. Available from: http://www.nlm.nih.gov/research/umls

[11] Medical Data Models. Available from: https://medical-data-models.org/ Accessed 2014 Dec 12.

[12] Follow Up Form. Available from: https://medical-data-models.org/forms/5232 Accessed 2014 Dec 12.

[13] Institute for Applied Quality Improvement and Research in Health Care. Available from: https://www.aqua-institut.de/en/home/index.html Accessed 2014 Dec 12.

[14] Microsoft Office: Access. Available from: http://office.microsoft.com/en-001/access Accessed 2014 Dec 12.

[15] Varghese J, Dugas M. Frequency Analysis of Medical Concepts in Clinical Trials and their Coverage in MeSH and SNOMED-CT. Methods of Information Medicine, 2014: 53(6), Epub ahead of print.

[16] Gangopadhyay A. Methods, Models and Computation for Medical Informatics. IGI Global. 2012 December; p 173.

Address for correspondence

Julian Varghese, Physician, M.sc. Computer Science, Institute of Medical Informatics, University of Muenster, Albert-Schweitzer-Campus 1, Gebäude A11,48149 Münster,

Germany. Email: julian.varghese@ukmuenster.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-842

vizHOME - A context-based home assessment: Preliminary implications for informatics

Gail R. Casper[a], Patricia Flatley Brennan[a,b], Jesse O. Perreault[b], Alex G. Marvin[b]

[a] *School of Nursing, University of Wisconsin, Madison, WI, USA*
[b] *College of Engineering, Industrial & Systems Engineering, University of Wisconsin, Madison, WI, USA*

Abstract

The rapid migration of health care from the institution to the home presents a plethora of consumer health technology options.. The fit of these technologies to the users' actual task performance and environment remains to be explored. In the vizHOME study, we set out to conduct in-depth analyses of health information management tasks conducted by individuals residing in 20 homes in the Midwestern United States who self-reported with diabetes. This paper will explore early results from five of the 13 assessments we have performed to-date. Early observations are described and implications for informatics are posited.

Keywords:

Self-management; Personal health technologies; Chronic illness; Task analysis; Human factors.

Introduction

The rapid migration of health care from the institution to the home presents a plethora of consumer health technology options. However, few of these are developed with an understanding of chronically ill individuals' health information needs, their abilities to meet these needs and how they use this information in managing their health and health care. Importantly, the place within which a person conducts specific tasks is rarely documented or explored. Yet there is growing evidence that context shapes health information management behaviors [1]. There is also growing evidence that better understanding of these self-management behaviors might lead to both clarification of the health information management challenges faced in the home and improved design of the technologies intended to be used there [2]. For example, Valdez et al. [3] propose a patient work model which involves attending to "the embeddedness of patients' health management in larger processes and contexts and prioritizing patients' perspectives on illness management".

The vizHOME project is designed to systematically determine how household context shapes self-health management including personal health information management (PHIM). PHIM encompasses a suite of cognitive and behavioral tasks that people undertake to accomplish their health goals. These include monitoring health states, recording symptoms, communicating with clinicians, as well as making sense of discharge summaries, health-related web sites and clinician-provided handouts. Often people undertake PHIM at home, and the home's physical features such as storage adequacy, lighting, privacy, and proximity of health information management tools influence the person's ability to recognize and complete the cognitive and behavioral tasks. Knowledge of which features of the household context shape PHIM can be used to better design technologies for PHIM. Our long-term goal is to improve individuals' self-management and health outcomes by accelerating the design and adoption of PHIM-supporting computer technologies that explicitly take into consideration features of the home context.

A key element of this study is in-depth assessment of the individual's self-management behaviors or self-care task performance as well as the devices they use in order to better understand the informatics issues in the design of tools and technologies used for these tasks. The focus of this paper is on the first phase of the vizHOME study, specifically on the exploration and documentation of the individual's performance of self-management tasks including such things as medication management and use of a monitoring device. A LiDAR scanner is used to capture the entire interior of participants' homes; these images are subsequently rendered for viewing the context of the homes in a 3D CAVE. A detailed discussion of the technology and processes used will be reported elsewhere.

The target population for this study is adults who report having been diagnosed with diabetes. Estimated to double globally from 171 million people in 2000 to 366 million people by 2030, diabetes is expanding both the amount of personal health technology many patients use regularly [4] and the personal health information management (PHIM) demands placed upon them. Individuals diagnosed with diabetes have a large burden for self-management, including self-monitoring of blood glucose, observing for complications, managing an array of ingestible and injectable medications and adhering to instructions from multiple health care providers. Each of these tasks have significant information management demands (recognition, recording, communicating, interpreting) and many technologies are being developed to assist patients to manage these tasks. Situating this first investigation of how household context influences PHIM in the homes of individuals with diabetes is valuable because much of the management of this disease occurs in the home, interspersed within their daily lives.

The task assessment is guided by SEIPS, the Systems Engineering Initiative for Patient Safety model [5]. This model posits that the work system (in this case, a *patient* work system), a structural concept, influences how processes occur, which in turn shape outcomes. The health care work system is defined as the integration of a person attempting one or more tasks using a given suite of tools and technologies in a specific physical environment influenced by the policies and procedures of a specific organization. We translate this to the home context, in which the person is the lay individual who is attempting self-management tasks for self or others using the available supplies, computer tools and resources in the informal household environment following the culture, family practices and other influences on intimate life at home. In this project we will focus on the work system in the home, with particular attention to aspects of the physical environment and personal health information management tasks.

Methods

Institutional Review Board approval was obtained for both project development activities as well as for the in home assessments from the University of Wisconsin-Madison Health Sciences IRB.

Sample

We partnered with The Survey of the Health of Wisconsin (SHOW) project to identify participants for this home task assessment. The SHOW has been conducting annual surveys since 2008, and at the beginning of 2012 had a growing cohort of over 2,500 Wisconsin adults. Eligible participants are men or women between 21and 74 years old who self-report being told that they have diabetes mellitus. They must be cognitively intact and able to read and write English. To enable in-home conduct of the interview, participants must live within a 100-mile radius of the project team's home institution in Central/South Central Wisconsin. Target enrollment is 20 participants for this home assessment phase of the study. We targeted five participants who live in one of each of the following home types: detached, semi-detached, multi-unit and mobile home. Twelve home assessments have been completed to date.

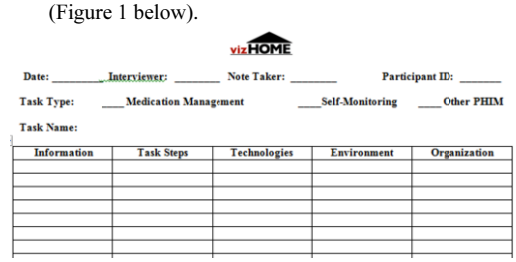
Data Collection Procedures

Interviews were conducted by study team members with a background in human factors engineering or nursing. A second person whose role is to take detailed notes is usually present as well. With participants' agreement, interviews were also audio-recorded for quality assurance purposes. After providing informed consent, participants respond to questions about their health concerns and self-monitoring and self-management practices. Interviews were conducted iteratively over three home visits using researcher-developed semi-structured interview guides based on the SEIPS Model. They were usually conducted solely with the participant unless the participant requested another household member be involved. Therefore, the participant is the primary source of information about the task performance. When the participant identified another household member as a better source of information about a specific process, that member was either included in the interview process or consulted at another time. A schematic map was created during a walk-through of the home during the first visit. This map served as a model that enabled us to anchor tasks or specific task steps to specific locations in the home. The participant was asked to identify key areas during the initial walkthrough and subsequent visits, and the interviewer requested the participant demonstrate specific tasks in the locations where they were actually performed.

The first visit most often included the interviewer and note taker. Subsequent visits (two and three, four if needed) involved both the interviewing team and the scanning team. Because people could not be in the area where scans were being done, the interviewers and scanners coordinated plans and activities.

The interviews and the intervening analysis proceeded as follows:

- Interview 1: introduction to the study and study team; obtain informed consent; questions about health and health information management tools, *health concerns inventory and health task inventory*.
- Intervening Analysis: *selection of three tasks for follow-up in Interview 2*. Individually and in pairs team members reviewed interview notes and audio recoding to discern three information-intensive tasks that are important to the individual. Responses were topically organized into three task types: medication management, self-monitoring and management of health information, noting specific indicators of all elements of the work system (information, task steps, tools and technology, environment and organization). Tasks were refined in a way to guide the scope of the in-depth exploration.
- Interview 2: *Generating rich descriptions of the selected three tasks.* One interviewer and the note taker returned to the home to undertake an in-depth exploration of each of the thre tasks. As the interview focused on the information intensive tasks undertaken by the participant, a grid was used to facilitate notetaking with a focus on information and the remaining four elments of the SEIPS model (Figure 1 below).


vizHOME

Date: _____ Interviewer: _____ Note Taker: _____ Participant ID: _____

Task Type: ___Medication Management ___Self-Monitoring ___Other PHIM

Task Name:

Information	Task Steps	Technologies	Environment	Organization

Figure 1 - Task Analysis Grid

- Intervening Analysis: Between interviews two and three, the interview team reflected on participant responses and created a transcription of the modified SEIPS grid as described in detail in the Analysis section below.
- Interview 3: *Respondent confirmation.* This interview was conducted to validate the interviewer's analysis and understanding of the task information and to obtain clarification when needed.
- Summative Analysis: Following the third interview, synthesis documents were created. These summary documents include task process summaries, a home persona, and an integrative case summary that addresses the tasks performed in the context of the home and technologies used.

Data for a cumulative visit log and background information was collected over the series of visits. Data included date and time of visits, location of the home (urban, suburban or rural), length of time in home, any perceptual or mobility limitations experienced by the key respondent, a list of others living in the home and presence of pets. Using a visual analog-type scale, the interviewer and note taker also rated the amount of clutter in the home. Differences in the two clutter ratings were discussed and resolved either by agreement or interpolation between scores.

Analysis

One task grid was created for each of the three tasks reported by the participant from each household. These were stored in computer files as pdf documents that could be annotated by team members.

The actual analysis of the raw data to create task grids for each task and each individual proceeded as follows.

Following the second in-home visit, the detailed notes from each task analysis were used to create a Task Analysis Grid for each task. The intent of the Task Analysis Grid was to create an "instruction manual" for the task *as the participant re-*

ports performing it, including the information (content and strategies) used so that the task could be replicated independently by anyone reading the grid. Task steps were arranged sequentially in the order the participant reported (or was observed) executing them. Efforts were made to discern task steps sequentially during the interviews. When this was not possible the analysis process involved imposing a logical task order response denoting information used, task step performed, technology or tool used, environment (i.e. location in the home), and organization were separated into the appropriate columns within the modified SEIPS grid. The interviewer created the grids using his or her own the notes as well as those written by the note taker in order to minimize errors and personal bias. If inconsistencies were encountered, the two home visitors discussed these differences and if a consensus could not be reached the topic was revisited in the third in-home visit.

After the Task Analysis Grids were completed to the extent possible, missing elements (evident as holes in the grid) were highlighted and questions structured to address these gaps. This missing information, as well as any other questions that emerge when constructing the grids, were compiled into an interview guide for the third in-home visit for the purpose of confirmation and clarification.

At the end of each household assessment we have an electronic file that includes a profile of each home, data about the house itself, descriptions of the patient work system, and descriptions of context (text, still photos and drawings). The data analysis process applied the Health@Home model [6] to begin to characterize information management practices by purpose and location. Additional analysis focused on determining whether there were differences in how similar tasks are done in different households. Finally, we reviewed home maps and images to facilitate reconstructing task activities across space.

Other analyses that will be undertaken, but are not reported here include: data cleaning and verification, descriptive statistics on demographics, background health information and household environments, and task processes.

Results

The five participants described here were systematically selected as representative and presenting a range of behaviors related to each task. A profile of each participant is presented. The findings are organized by task: medication management, self-monitoring and information management.

Demographics

The five participants range in age from early 30s to early 70s. They are diverse in terms of racial and ethnic background. Two participants live alone, two live with a spouse and one lives with a spouse and pre-school children. One participant resides in semi-detached home, two in multi-unit and two in a detached home, thus they represent four of the five home types.

Participant Profiles

Participant A lives alone in a one-bedroom apartment in an urban area. She has several chronic illnesses in addition to diabetes. She is solely responsible for her health and healthcare. She has a primary care provider, a diabetes specialist and a case manager.

Participant B is retired and lives with his wife in a semi-detached suburban home. In addition to having type 2 diabetes, he has several chronic health conditions, most related to his heart. He regularly (one to four times per year) sees a primary care provider, a cardiologist, an electrocardiologist and an endocrinologist.

Participant C has insulin dependent diabetes. He lives with his wife in a detached urban home and is employed full time. He has hypertension and elevated cholesterol.

Participant D is a retired woman who lives alone in a two-bedroom apartment in an urban area. She self-reports as obese and has elevated cholesterol, chronic back and hip pain.

Participant E is a married man who lives with his wife and children in a detached suburban home. He works full-time outside of the home. He has several chronic conditions including insulin dependent diabetes, hypertension and elevated chrolesterol.

Task: Medication Management

Participant A*:* In addition to the prescribed 12 pills taken daily, Participant A uses eye drops at least daily, takes *as-needed* medications for pain and airway management and self-administers a weekly injection. She uses a weekly pill organizer to manage her twice-daily medications. She describes her medication taking as a routine for which she does not require cues. Participant A keeps her pill organizer on a bedside table and an injectable is stored in the refrigerator. She "rarely" forgets to take prescribed doses, but knows what to do when she does forget. Her source of medication information is primarily her primary health care provider. She also receives drug information from the pharmacy which she carefully reads and then immediately discards.

Participant B takes nine medications daily, divided between the morning, dinner time and bedtime. He takes his medications from the bottles in which they are dispensed. All but one medication bottle are kept on the bureau in his bedroom; each morning he counts out the day's pills and lays them on the bureau. At some point during the day he moves the pills to be taken in the evening into the kitchen and puts them on the counter. The single medication taken with the evening meal is kept on the dining table for convenience and to serve as a reminder to take it. He is meticulous in following his prescribed medication regimen and almost never forgets a dose. He maintains a hand-written list of current medications and revises it whenever changes occur. He keeps pharmacy inserts for current medications organized in a folder in his bedroom and states he would refer to them if he had any questions about his medications.

Participant C takes five different medications daily, including insulin in the morning and evening. Each day he takes his morning pills out of bottles and puts them in his pocket to take at work. He also takes a bottle of insulin and a syringe to work for his morning dose. He takes his evening insulin and oral meds when he gets home from work. His medications are stored in a bathroom cabinet that he shares with his wife. He checks the bottle label each time he takes a dose to ensure he is taking the right medication. He "sometimes" forgets to take medications to work with him; when this happens he resumes regular dosing the next day. He relies on his pharmacist to relate relevant medication information and discards informational inserts when he receives them.

Participant D takes 14 pills per day for several chronic conditions including "borderline" diabetes, hypertension, hypercholesterolemia, heart disease and arthritis. Her health care is delivered through a community-based clinic that includes a pharmacy, a senior center and social services as well. A pill organizer is filled by the clinic pharmacy and delivered to her on a weekly basis due to difficulty in maintaining correct dosing. She keeps this organizer on a

table in her bedroom. If she is going out for the day or plans to stay in her living room, she transfers single doses into a small pillbox to keep with her. Routines around waking up and retiring for the night serve as reminders to take her medications at those times; she uses specific shows on television as cues to taking daytime meds. She does not receive pharmacy inserts with her medications; the organizer has labels with the names of the medications it contains on the bottom. She stated that she "would not understand the detailed information anyway". She reports that she sometimes forgets to take a dose and will take it when she remembers "if it isn't too late" or will just skip it.

Participant E has oral medications that he takes in the morning and evening. He also takes scheduled doses of mixed insulin in the morning and before his evening meal. In addition, he has "sliding scale" regular insulin prescribed four times per day, but he rarely checks his blood glucose more than one to two times per day because he is "busy". His oral medications are stored in cabinet in the master bathroom; pill bottles are organized on a shelf from right to left according to time taken. He takes the morning oral medications, one at a time, in the bathroom. The insulin pens are stored with his glucometer in a drawer in the kitchen. He stands at the kitchen counter above the drawer to use his glucometer and inject the insulin. He might read medication information when a new med is started otherwise discards information when he receives it.

Task: Self-monitoring

Three of the five participants described above received a glucometer to monitor their blood sugar at varying frequencies from daily to several times per day. One participant adhered to the prescribed frequency while others used the glucometer 1-3 times per day less than expected. Some occasionally manually logged the glucometer readings, usually to share with their health care provider. While all had a glucometer with a history feature, none reviewed readings in the history. One participant did sync the glucometer to his smart phone and sometimes reviewed readings in it. All reported knowing what range their readings should be in and what to do if the readings were not in range.

Table 1 - Self-monitoring

Participant	A	B	C	D	E
Glucometer Prescribed	yes	no	yes	no	yes
Frequency	1+	*heart	2-3		1-2
Storage Location	bed-room	phone	back-pack		kitchen
Use Location	bed	bed-room	work, home		kitchen
Records Results	some	digital	no		no
Takes to MD appts	yes paper	dials in	synchs to phone		yes glu-cometer

Remarkably, they attributed such information to learning from their health care providers and all was committed to memory. Table 1 summarizes the particicpant's self-monitoring of blood glucose level. Note that Participant B did not have a glucometer, but did regularly monitor his cardiac status. He used telemonitoring to provide data to his cardiologist and has done so since he had an implanted defibrillator inserted years ago.

Task: Information Management

Participants reported a variety of policies and approaches to information management. For some, information was defined as a reminder or cue to begin a task; for others it was in the form of care instructions conveyed by a health care provider.

Table 2 - Information Management

Partici-pant	A	B	C	D	E
Format	paper	paper	paper	paper	paper
Seeks	no	yes	no	no	yes
Reads	yes	yes	about device	occa-sionally	scans
Shares	no	w/wife	no	no	no
Disposi-tion	keeps new	keeps new	discards	dis-cards	dis-cards
Storage Location	bed-room basket	bed-room bureau	phone, if any	n/a	n/a

For still others it was paper-based and included self-care recommendations, manual-type literature, general health and disease-oriented publications or pharmacy-inserts. Remarkably, most reported having a personal health record (PHR) but regardless of their information seeking and management behaviors, none accessed the PHR. A summary of select information management strategies is contained in Table 2.

We noted a single commonality across tasks and participants – each had a stable set of strategies which they applied to the PHIM tasks they performed. In spite of this, we noted a high level of unique approaches in terms of the behaviors within tasks of similar types. In addition it was often difficult to determine discrete beginnings and ends of the task. Finally because all tasks are influenced by the fit within the person's life, there is a lack of a "gold standard" for performance.

Similarity/Difference Analysis

Generating design recommendations requires first determining the extent of similarities across people within tasks. Previous methods [7] were not applicable to this analysis. We undertook the following analyses in order to detect similarities and differences within tasks and across participants.

All tasks were first grouped by high-level similarity. For example, every participant taking daily medications was documented. Many participants self-monitored and treated blood sugar levels; this was also documented. There were several other groups of tasks that were performed by smaller numbers of participants. A group was defined as at least two participants performed a given task.

Two analyses were conducted on the data from the Task Analysis Grids: 1) location analysis and 2) task step similarity analysis. For each analysis a color-based coding key was generated. For the location analysis each different type of room or location where task steps occurred was represented by a different color (e.g. kitchen, bedroom, living room, closet, etc.). The coding key was limited to areas where task steps actually occurred in the study; it was not an exhaustive list of all possible locations in the homes that could be used. Multiples of the same type of room were documented when appropriate (e.g. master bedroom vs. other bedroom). For the task step similarity analysis a similar color-based coding key was created with high-level designations for the type of task step recorded (e.g. administering medication, documenting, assessing personal status, etc). This list was also limited to observed behaviors and not exhaustive of all possible behaviors. Selected columns of the Task Analysis Grids (i.e. Information, Task Steps and

Environment) for the tasks in each group were arranged side by side in a spreadsheet to allow for comparison. The two analyses were performed separately on transcribed copies of the grouped tasks.

In the location analysis the cells in the Environment columns of the Task Analysis Grids were coded by the location in the home where each task step occurred. When a task step involved transition between different locations or when it might be done in different locations the cell was marked with both colors.

In the task step similarity analysis the cells in the Task Step columns were coded by the type of task step that occurred. In similar fashion to the location analysis, when a task step spanned more than one type of behavior the cell was marked with multiple colors. In this analysis, some cells were left uncoded if the behaviors were not sufficiently close to one of the coded types or if they were redundant with other task steps. As identified earlier, when analyses were conducted at a relatively granular level, more differences than similarities were detected.

Discussion

The design and implementation of supportive technologies continues to proliferate. We learned from these five participants that while the tools may perform satisfactorily in usability assessment, they may not meet the needs of many users. An example is the growth and increased availability of personal health portals. Although many of the vizHOME participants had access to some type of personal health record (PHR), few if any used it. Although the scope of the study does not include exploration of non-use, we anecdotally learned there may be several underlying reasons related to access (technology), usefulness, or demand.

Another reported reason for non-use of tools aimed at supporting self-care is the level of the content. One participant reported she did not read or keep medication information from the pharmacy because it was "like reading a foreign language". Given this admission, one should consider the importance of providing information that consumers can understand and apply in self-monitoring and self-management. Translating such information into a vernacular that is readily understood by the consumer may be needed.

We also found that most of the vizHOME participants had access to a computer or had some type of internet access. Despite the capacity, they tended to not use not use electronic methods for health information access, interpretation or storage. With the increase in computer generated cues such as pushing information or reminding patients about health care appointments, consideration should be given to translating such cues to commonly used household items such as a calendar.

Conclusions

Health information management task performance demonstrated more differences than similarities than we originally anticipated. These differences present additional challenges to those who design tools and technologies for self-care. Indepth study of people as they manage their illness or health at home will continue; we expect that this will generate a better understanding as well as provide guidance for the design of self-care technologies.

Acknowledgements

This project was supported by grant number R01HS022548 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

The authors acknowledge Dr. Pascale Carayon and Dr. Peter Hoonakker for their contributions to the design of home assessments and analysis of interviews and observations. We also thank the members of the Living Environments Lab.

References

[1] Provencher V, Demers L, Gelinas I, and Giroux F. Cooking task assessment in frail older adults: Who performed better at home and in the clinic? Scan J of Occup Ther 2013: 20(5): 374-83.

[2] National Research Council. Consumer Health Information Technology in the Home: A guide for human factors design considerations. Washington DC: the National Academies Press, 2011.

[3] Valdez RS, Holden RJ, Novak LL, and Veinot TC. Transforming consumer health informatics through a patient work framework: Connecting patients to context. JAMIA online. doi: 10.1136/amiajnl-2014-002826.

[4] Wild S, Roglic G, Green A, Sicree R, and King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004: 27(5): 1047-53.

[5] Carayon P. Hundt AS, Karsh BT, Gurses AP, Alvarado CJ, Smith M, and Brennan PF. Work system design ffor patient safety: The SEIPS model. Qual & Safety in Health Care 2006: 15(SI): i50-i58.

[6] Moen A, and Brennan PF. Health@Home: The work of Health Information Management in the household – implications for Consumer Health Informatics. JAMIA 2005: 12(6): 648-56.

[7] Ozkaynak M, and Brennan PF. Revisiting sociotechnical systems in a case of unreported use of health information exchange system in three hospital emergency departments. J of Eval Clin Pract. doi:10,1111/j.1365-2753.2012.01837.x.

Address for correspondence

Gail R. Casper RN, PhD, Senior Scientist
University of Wisconsin-Madison
School of Nursing
701 Highland Avenue
Madison, WI 53705
gcasper@wisc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-847

The Use of Applications in Distance Education Specialization Course as a Support Tool for Students Living in Remote Areas Without Internet

Ana Emília F. Oliveira [a], Rômulo M. França [a], Eurides F. Castro Júnior [a], Deborah C. L. Baesse [a], Mariana F. L. Maia [a], Elza B. Ferreira [a]

[a] Open University of Brazilian National Health System (UNA-SUS), Brazil

Abstract

The world is experiencing the popularization of mobile devices. This was made possible by the increasing technological advances and the advent of the Internet as a communication and information tool. These facts demonstrate that the development of applications compatible with such devices is an effective way to provide content to diverse audiences. In the educational field, these devices can be seen as technological support artifacts for distance education, serving as strategy for continuous and permanent education for health professionals. The Open University of Brazilian National Health System (UNA-SUS) offers distance learning courses, including specializating on free access. In order to increase the public reach, UNA-SUS developed mobile applications as supporting material for students. These applications can be accessed in offline mode, increasing the accessibility and therefore, improving the efficiency of the material. The 28 applications developed with responsive online books format currently reached the milestone of over 6,000 downloads. This number shows the positive acceptance of the format used, accentuated by the ease of having material downloaded from the device, not requiring the user to be connected to access content.

Keywords:

Mobile Applications; Distance Education; Mobile Device; Permanent Education in Health.

Introduction

The combination of the Internet and mobile devices, such as smartphones and tablets, tools for communicating and spreading information has been on the rise. The popularization of mobile devices has been considered by many as the technological revolution with the greatest impact in recent times. The impact is not only because it allows more convenience in accessing information and communication, but also due to the characterisctics of such devices considered pocket computers, which allows access to millions of applications. According to the IDC (International Data Corporation) in 2012 [1], more than 40 billion apps have been downloaded on smartphones, and it is predicted that this number will reach 300 billion in 2016 [2,3]. Therefore, it is clear that developing applications whose format is compatible with mobile devices has become an effective way to provide certain content to specific target-audiences [4].

Information technology (IT) applied to education has transformed school environment and generated innovations in learning. The trend in teaching-learning models is to try to overcome time and space boundaries and reach students regardless of where they are or when they are available. Although this claim has not been fully achieved yet, distance education has occupied a significant space among the pedagogical tools used [5].

Distance education (DE) is a technological two-way communication system that may be considered as an alternative to the traditional classroom teacher-student interaction model and allows learning without the need of those involved in the process to be present in the same location interacting at the same time [5,6]. Therefore, despite the physical distance between teachers and students, distance education makes use of ways and technologies in order not to compromise the process of transmission of knowledge. In this case, mobile devices can be seen as a supportive technological artifact, so that distance education reaches the virtual integration of learning vectors.

If we pay attention to health professionals and to the difficulties faced by them, we realize that, more than any other, they need to constantly update their technical, technological and social skills, while respecting ethical principles governing their conduct. For these workers, distance education would be a strategy for continuous and permanent education, given the numerous technological advances and pedagogical innovations in education [7].

UNA-SUS is a project from the Brazilian Ministry of Health developed in partnership with the Work and Health Management Secretariat - SGTES in order to offer Brazilian National Health System (SUS) workers democratic, flexible and quality education, respecting employee's working time and their place of residence. Seeking to improve the qualifications of health workers through Distance Education, the Open University of Brazilian National Health System (UNA-SUS), in partnership with the Federal University of Maranhão (UFMA), offers specialization, training and extension courses with free access, with the goal of providing training in strategic areas of health care service.

The courses offered by UNA-SUS reach a target-audience with the following characteristics: residents living in of remote areas limited Internet access, lacking of time to study and having intense days of work. These factors would hardly allow them to attend traditional face-to-face education. If we merge the increasing demand for distance learning courses to the popularization of mobile communication devices, we have the ideal solution for the development of innovative teaching methods. Following this reasoning, UNA-SUS/UFMA is a pioneer within Brazil in developing applications for mobile devices as background material for students. These applications have technological capabilities allowing offline mode (no Internet access), increasing the material's accessibility and efficiency.

Materials and Methods

Territorial Reality

UNA-SUS/UFMA is located in the state of Maranhão, which has 217 municipalities, and a current population estimated at 6,850,884 inhabitants [8]. Since Internet access is limited in areas far away from metropolitan region, the state's vast territorial field represents a challenge for the attainment of distance education initiatives such as UNA-SUS/UFMA. Figures of 2010 Census show that Maranhão is the last in the list of federal units with computers connected to the Internet, with the capital São Luís having the highest access rate [9]. Moreover, according to a recent Teleco survey – the biggest information portal in Brazil's telecommunications sector – out of the 217 municipalities of Maranhão, 39% do not have 3G Internet coverage [10]. However, it is exactly in these regions where the most urgent demands for training are located, especially for health professionals, the very target-audience of UNA-SUS/UFMA.

Distance education has been considered an important tool by the Brazilian National Policy of Permanent Education in Health and has been well accepted by its participants. However the difficulties concerning the access to Internet have still been a limitation to the growth and success of these initiatives [11].

Theoretical Basis for Applications Development

According to research from Business Insider five years ago, much of Internet accesses from desktop computers. In 2013, 22% of the population owned a smartphone, an increase of approximately 1.3 million in the number of smartphones since 2009 [12].

Given the prevalence of smartphones with Internet access, distance learning must adapt its educational tools in order to reach those with mobile devices. Since it is not feasible to build specific websites or applications for each platform and device that a user can use, adjustments must be made so that applications work correctly in any variety of mobile equipments that one can use to access educational material [13].

The Responsive Web Design (RWD) was chosen as a solution that enables adaptation to user behavior and to the environment used to access the information. It also takes into account the platform, the resolution and screen orientation. Responsive Web Design includes techniques and technologies that are adjusted to make a single application run on a variety of devices in the most utility way possible [14].

Responsive Design aims to attend this kind of demand by fulfilling the following requirements [15]:

- Adapting layouts to different screen sizes;
- Resizing pictures according to screen resolution;
- Supplying images optimized for the aspect ratio of mobile devices;
- Simplying page elements to mobile versions;
- Hiding non-essential elements on smaller screens;
- Providing buttons and links with large clickable areas for mobile devices;
- Detection of mobile devices' specific features such as screen geolocation and orientation.

Besides Responsive Design's aspects, to obtain quality material capable of providing a motivational and pleasant educational experience, it is essential to develop it on pedagogical notions basis about educational interfaces that are favorable to learning and knowledge assimilation.

An attractive interface with appropriate interaction devices has a positive impact on the software's usability, acceptance and learning promotion capacity. Therefore, the software's design and its interface should be aligned with appropriate pedagogical principles, providing characteristics such as content and interaction presentation format up to the amount of information presented [16].

Hence, applications created and made available by UNA-SUS/UFMA are developed based on the learning objects concept, aimed at being integrated into different hardware platforms, including mobile devices, in addition to enabling access without the requeriment for Internet connection. The design of these learning objects prioritizes content, student interest and learning theories [17].

Learning objects are educational materials for pedagogical purposes that serve to support teaching-learning process. They are built in small groups or blocks with which learning content is structured. Their use is more efficient when they are cataloged on metadata and stored in a repository integrated to a learning management system, so that, in case of replacement or platform upgrade, there is no need for redesign [18].

Learning objects cataloguing offers other benefits beyond reusability, such as [17]:

- Accessibility: ability to access educational resources in a remote location and use them in many other places;
- Interoperability: possibility of using components developed in one location, with a set of tools or platforms, in other places with other tools and platforms;
- Durability: ability to continue using educational resources when their technology base changes without having them redesigned or reprogrammed.

Academic software from UNA-SUS/UFMA provide students with educational content for their courses in a digital book format, ensuring interactivity, a key aspect to learning process. Hypertexts and images, which meet visual adaptation criterias set by Cognitive Theory of Multimedia Learning, have become attractive and understandable to students and enabled the possibility of 'navigation' the way they want, respecting their time of learning [18].

Cognitive Theory of Multimedia Learning states that, once combined, multimedia narration and graphic images produce verbal and visual mental representations that integrate with prior knowledge to build new knowledge. This way, words and corresponding pictures must be close together, since they facilitate construction of related links between them [19]. Therefore, the use of images to illustrate concepts is essential in educational content.

Another important aspect to promote meaningful learning in a virtual environment is to design activities that reduce cognitive load, free memory capacity for deep cognitive processing during learning, increase students interest, and encourage them to use that freed memory for deep processing during learning. In others words, the interest can be stimulated by simply presenting the material in a visually appealing way, accompanied by text or animated and pleasant narration [20]. Therefore, a major concern of UNA-SUS/UFMA during the creation of its applications is to make adaptive images, so they do not suffer changes that could make them dysfunctional in the transmission of available content.

Educational Planning

Within the production of educational resources, UNA-SUS/UFMA provides content planning that will be used in its applications: defining all supporting materials to be used in the module and, later on, didactic planning of these modules. In the planning of supporting materials, definitions of all the tools that will be used in the module, such as videos, support texts and description of assessment activities, always trying to choose these materials carefully, and evaluating the possibility of their access in the application [21].

Didactic planning takes into account the course's or module's workload, target-audience and learning objectives. In the planning of each unit, learning objectives are defined, description of the discussed content is created, designing the presentiaton of the module and suggestions of learning evaluation are made. Based on this didactic planning, modules are built from the content and activities definition text to the process of Instructional Design.

For applications that work in offline mode, it is taken into account that the purpose of their construction is to facilitate access to course content developed by UNA-SUS/UFMA, especially for students who work in locations with limited access to Internet services. Therefore, the option for not providing online supporting materials and assessment learning activities is purposeful, so that didactic and modular planning is architected with the necessary adaptations for this technology [21].

Application Creation Process

The application development process took place with participation of two sectors of the institution: Institutional Design sector (ID) and Design and Information Technology Center (ITC), each one having a different responsibility in applications development. Institutional Design sector is responsible for developing all educational content, prepared by professors responsible for creating online content, which will be available to the student. The Design sector is responsible for developing the entire graphic part of online books such as icons, images and color palette. Finally, ITC has the responsibility of developing digital book in form of mobile applications and publish it in Virtual Learning Environment (VLE).

Applied Technology

Applications created by UNA-SUS/UFMA are developed in HTML5 (HyperText Markup Language 5) CSS3 (Cascading Style Sheets 3) and JavaScript. In addition, mobile versions of digital books are generated as an application for Android and iOS platforms, using Phonegap/Apache Cordova technology.

HTML5 is suitable for the development of responsive interfaces. It offers advantages that HTML4 and XHTML languages do not, such as the fact that most mobile devices have browsers that support HTML5 and CSS3. This type of approach in cross-platform development frameworks for mobile devices is intended to provide better user experience, with greater flexibility and media insertion possibilities [22].

In addition, applications developed in this language are optimized for smaller screens. The device's operating system has no control over their content or functionality, but developers can make changes at any time with immediate effect for users [22].

Application Presentation

All applications for mobile devices are available in both platforms of Google's store (Play Store) and Apple's store (App Store), and can be found through the search tool, using the words UNA-SUS UFMA (https://play.google.com/store/search?q=una-sus%20ufma). Currently UNA-SUS / UFMA has its whole application in English, and its content is dedicated to Postpartum Depression (https://play.google.com/store/apps/details?id=com.unasus.appdpp).

Currently, 28 modules were already developed and made available in the format of responsive online books by UNA-SUS/UFMA, enabling the mark of more than 6,000 downloads until this moment. This tells us that there is a positive acceptance of format used, accentuated by the ease of having the material downloaded from the device, not requiring users to be connected whenever they need to access content.

Figure 1 illustrates the presentation screen of the online book, visualized from both desktop and mobile device screens. All the elements of each page (navigation bar with arrows, navigation between pages, the tool to increase the font size) can be well accessed through the application, because the buttons were designed to be compatible to human finger size.

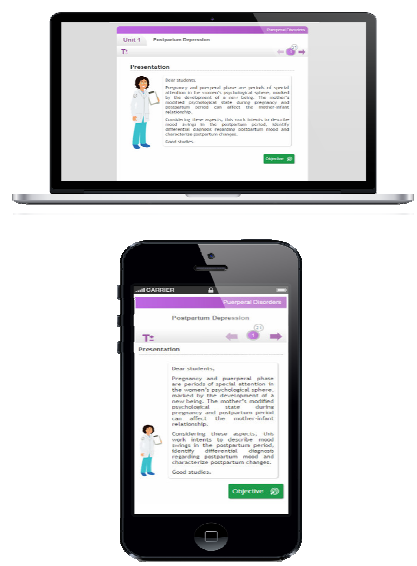


Figure 1 – Screen Presentation Mobile and Desktop

Figure 2 shows the contents view of a unit of digital book about Postpartum Depression (PPD) in the application.

By clicking on each gray button (Baby Blues, Postpartum Depression and Puerperal Psychosis) the text box below the woman's figure changes, providing detailed information concerning the gray button clicked on. This way, there is a space on the screen is saved, allowing a significant amount of content on a page without it being polluted with too much text, which would make the reading difficult and tiresome, especially on the small screens of mobile devices. All techniques used to build the applications are designed to make each online book page less visually polluted, while providing important informations on each topic, which means that there is no reduction in the content, but use of technological strategies for dynamic and smart use of the area in each page.

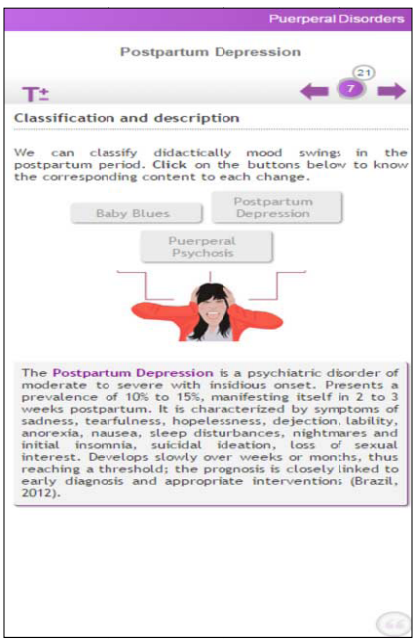


Figure 2 - Screen of a Content from a Digital Book in the App

Given the experience of UNA-SUS/UFMA in applications for mobile devices, some points can be highlighted as positive in the use of this material as a resource for students living in remote areas without Internet access, such as [21]:

Mobility, that refers to the fact students can load applications modules everywhere, since their mobile devices remain connected throughout the day due to its multiple features and possibitilies for locomotionincreases the mobility of the application. Ease of access is directly related to offline availability: students only need to install the application once on their device, in others words, only one Internet access is absolutely mandatory. From this, the application is available on your mobile device and can be accessed even without Internet connection. This attribute is one of the most relevant to UNA-SUS/UFMA, since most of its students live far away from the state's capital and have difficulties accessing Internet. Thus, studies are not interrupted if there are any connection problems: digital books' content is available at any time on students' device.

If there are updates for mobile application regarding new content or corrections, automatic updates are performed. In such cases, there is no need for students to intervene since this work is done by the web server of application's store. The cross-platform interoperability is characterized by the fact that digital books are developed once in HTML5 and adapted as open applications through PhoneGap. The specific implementation is not required for each mobile device platform (Android and iOS). Another positive and differential point regarding UNA-SUS/UFMA's applications can be found in the experience made available to the user. This is because content, images and animations and others features offered to students in the digital book of Virtual Learning Environment on their computer are available in the same way on the corresponding application for mobile devices. This means that students will not be harmed in relation to content or in any pedagogical sense for studying on screens with lower resolutions [21].

Conclusion

The popularization of mobile devices with Internet access has created an unique opportunity for Distance Learning, since it allows learning to happen anywhere, anytime, without relying on a physical connection line and with an equipment whose cost is lower than computers in general.

In this paper, we showed how UNA-SUS/UFMA took advantage of this new perspective on distance education to reach students living in remote areas where Internet access is limited, important obstacle to the completion of courses in this modality of education. Using the HTML5 markup language and the techniques of Responsive Web Design it was possíble to provide online books - the main source of content used in courses offered by the institution - on mobile devices.

The major advantage of this application is that it requires only Internet access once, the one in which the user downloads the app. After that the online book is available on users' devices without the need for Internet access. Therefore, although the students of the institution face problems accessing the Internet, the course content might be always available for consultation and studies.

References

[1] International Data Corporation (IDC). Worldwide Business use Smartphone 2013-2017: forecast and analysis. Framingham; 2013. Available at: http://www.idc.com/.

[2] Camacho AL. Análise das publicações nacionais sobre educação à distância na enfermagem. *Rev Bras Enferm*. 2009: jun./ago.; 62(4): 588-93. Available at: http://www.scielo.br/scielo.php?pid=S0034-71672009000400016&script=sci_arttext

[3] Tibes CS, Dias JD, Zem-Mascarenhas SH. Mobile applications developed for the health sector in Brazil: an integrative literature review. *Rev Min Enferm*. 2014: abr/jun.; 18(2) 479-486. Available at: http://www.reme.org.br/artigo/detalhes/940

[4] Rosa D, Silva TL. Adaptação de interfaces para dispositivos móveis com HTMLS5. Anais do EATI Frederico Wesphalen; 2013: nov. 249-252. Available at: < http://www.eati.info/eati/2013/assets/anais/artigo249.pdf>.

[5] Costa MLT. Ambientes Virtuais de Aprendizagem na educação a distância na pós-modernidade. *Rev Eletrônica da Faced Dialógica.* 2007. 1(20). Available at: http://www.historia.uff.br/estadoepoder/6snepc/GT7/GT7-SILVANIA.pdf

[6] Bévort E, Belloni ML. Mídia-Educação: conceitos, história e perspectivas. *Educ. Soc.* 2009: set./dez; 30(9):1081-1102. Available at: http://www.scielo.br/scielo.php?pid=S0101-73302009000400008&script=sci_arttext

[7] Oliveira, MN. Educação à Distância como estratégia para a educação permanente em saúde: possibilidades e desafios. *Rev Bras Enferm* 2007: set./out.; 60(5):585-9. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672007000500019

[8] Instituto Brasileiro de Geografia e Estatística [homepage na internet]. Progressão do índice populacional para 2014. [access at 3 dez 2014]. Available at: http://www.ibge.gov.br.

[9] Instituto Brasileiro de Geografia e Estatística [homepage na internet]. Resultados do censo 2010. IBGE, 2010. Available at: http://www.ibge.gov.br.

[10] Teleco – Inteligência em Telecomunicações. virtual community of professionals [homepage on internet]. Available at: http://www.teleco.com.br/3g_cobertura.asp

[11] Fuzissaki MA, Clapis MJ, & Bastos MAR. Consolidation of the national policy of permanent education: an integrative review. *Journal of Nursing UFPE*, 2014 [JNUOL/DOI: 10.5205/01012007], 8(4), 1011-1020. Available at: http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/viewArticle/3785

[12]Business Insider [homepage na internet]. One in every 5 people in the world own a smartphone: one in every 17 own a tablet. Available at: http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10#ixzz3Lh950JDM

[13] Alban A, De Marchi ACB, Scortegagna SA, Leguisamo CP. Ampliando a usabilidade de interfaces web para idosos em dispositivos móveis: uma proposta utilizando design responsivo. CINTED-UFRGS Novas Tecnologias na Educação 2012: dez; 10 (3). Available at: http://seer.ufrgs.br/renote/article/view/36404

[14] Prostt, ME. Interface Web utilizando design responsivo: um estudo de caso aplicado a smartphones, tablets, computadores e televisores. Curitiba: Universidade Tecnológica Federal do Paraná; 2013 Available at: http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/2513/1/CT_TECJAVMOV_I_2012_12.pdf

[15] Doyle M. Responsive Web design demystified. Elated, 2014: nov. Available at: http://www.elated.com/articles/responsive-web-design-demystified/

[16] Reategui, E. Interfaces para softwares educativos. *Renote.* 2007: 5. Available at: http://www.cinted.ufrgs.br/ciclo9/artigos/1bEliseo.pdf

[17] Tarouco LMR, Grando ARCS, Roland L, Roland P. Alfabetização visual para a produção de objetos educacionais. *Renote* [serial online], 2003: set; 1: 9. Available at: http://www.cinted.ufrgs.br/renote/set2003/artigos/artigo_anita.pdf

[18] Tarouco MR, Fabre CJM, Grando A RS, Konrath M LP. Objetos de aprendizagem para M-Learning. Congresso Nacional de Tecnologia da Informação e Comunicação; 2004. Florianópolis. Available at: http://www.cinted.ufrgs.br/CESTA/objetosdeaprendizagem_sucesu.pdf

[19] Mayer RE, Moreno RA. Cognitive theory of multimedia learning: implications for design principles. 2007. Available at: https://gustavus.edu/education/courses/edu241/mmtheory.pdf

[20] Sorden, SD. A Cognitive approach to Instructional Design for Multimedia Learning. Northern Arizona University Flagstaff, AZ, USA. *Informing Science Journal* 2005: 8. Available at: http://inform.nu/Articles/Vol8/v8p263-279Sorden34.pdf

[21] Oliveira AE, França RM, Pinho JO, Castro Júnior, EF, Baesse, DC. Aplicativo aberto para dispositivos móveis e sua aplicabilidade na atenção básica: a experiência da UNA-SUS/UFMA. In: Gusmão CM organizadora. Relatos do uso de Tecnologias Educacionais na Educação permanente de profissionais da saúde no Sistema Universidade Aberta do SUS. Recife: UFPE; 2014. Available at: http://www.unasus.gov.br/sites/default/files/livro_relatos_una-sus_2014.pdf

[22] Silva MM, Santos MTP. Os Paradigmas de desenvolvimento de aplicativos para aparelhos celulares. *T.I.S.* 2014: maio./ago; 3(2): 1 62-1 70. Available at: http://revistatis.dc.ufscar.br/index.php/revista/article/view/86/80

Address for correspondence

Ana Emília F. Oliveira

Rua Viana Vaz, n° 41, Centro, CEP: 65020-660. São Luís – MA, Brasil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-852

An Architecture for Continuous Data Quality Monitoring in Medical Centers

Gregor Endler[a], Peter K. Schwab[a], Andreas M. Wahl[a], Johannes Tenschert[a], Richard Lenz[a]

[a] *Computer Science 6 (Data Management), Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany*

Abstract

In the medical domain, data quality is very important. Since requirements and data change frequently, continuous and sustainable monitoring and improvement of data quality is necessary. Working together with managers of medical centers, we developed an architecture for a data quality monitoring system. The architecture enables domain experts to adapt the system during runtime to match their specifications using a built-in rule system. It also allows arbitrarily complex analyses to be integrated into the monitoring cycle. We evaluate our architecture by matching its components to the well-known data quality methodology TDQM.

Keywords

Medical Informatics; Data Quality; Quality Control; Quality Improvement; Data Collection; Databases, Factual; Group Practice.

Introduction

Data quality is an important concern in data integration scenarios. Integration and data quality influence each other bilaterally. Many data quality issues only become evident if several data sources are available, and integration can benefit from high data quality [1].

In the medical domain, one such integration scenario is created by medical practitioners affiliating into group practices or medical centers [2]. They do so for various reasons, among them financial benefits or better collaborative treatment of patients. To reap these benefits, organization wide planning is required. This in turn makes it necessary to create a general view over the data and processes of every participant's practice. This information is found in the respective local patient-data management systems and databases and may exhibit various deficiencies, like missing entries or wrong values. An integrated central database is fed new and possibly erroneous data from the center's locations continuously.

It is thus not enough to do a single pass of data cleaning over local databases. Indeed, it may even be impossible to validate data at the local databases since practitioners may be unwilling to relinquish control of their system or may be unaware of newly established quality considerations. Thus, central continuous monitoring and improvement of data quality is necessary to support the center. Frequent changes also induce significant software aging [3] in any system dealing with data quality. Data quality systems must be able to evolve with newly arising requirements, especially in the health sector. Medical or legal changes must be implementable by domain experts, since these changes may happen too short-term and too often to allow assigning a programmer for substantial code redesign [4].

Contribution

In this paper, we describe an architecture for continuous data quality monitoring. We start out by giving an overview over the project context, and describe our approach and the resulting architecture. To evaluate the architecture's usefulness, we show how it supports data quality methodology. In conclusion, we briefly discuss the ongoing implementation of a prototype.

MEDITALK

The research project MEDITALK is an example of an integration scenario in the medical domain [5]. A so called practice manager is responsible for controlling a medical center. This role necessitates an overview over the data of all the centers' locations. All local practitioners' data is integrated into a central database in a standard Extract-Transfrom-Load (ETL) process [6]. This solution is already in place at three medical centers, together encompassing 50,000 patients, 80 practitioners, and 3 practice managers. Since new data arises at the local practices continuously, ETL is repeated at set intervals. Copies of all local data arrive at daily intervals in the central database. The practice manager interacts with the system through a controlling application, which essentially provides a predefined dashboard of information about the centers' locations and their data.

Data Quality

Data quality is often generically defined as "fitness for use" [7]. This means that for data to be of high utility, it has to conform to requirements according to its application. We divide data quality considerations into two broad groups. Checking their fulfillment may be conceptionally easy, like verifying the length of a string. It may also be more involved, for instance requiring complex data mining or statistical models. Obviously, the boundary between these two difficulty groups is fuzzy - deciding which group a requirement belongs to is a domain expert's responsibility. Within both groups, requirements are further (sometimes implicitly) classified into the usual data quality dimensions [1,7]. We make the assumption that both groups are important to continuous data quality monitoring in medical centers.

Methods

In a first step, we conducted a literature review focusing on data quality both in general as well as specifically in healthcare. To familiarize ourselves with the domain, we afterwards conducted interviews with the three practice managers involved in MEDITALK. We asked questions about the data quality dimensions most mentioned in literature, their impact on the centers' work and about how detected quality problems had been addressed. We worked with the developers of the integration environment of MEDITALK, and designed an architecture to monitor data quality on top of the existing software.

Results

Literature Review

Many methodologies for data quality improvement have been proposed (see [8] for a survey). A commonality between these is that they offer a bird's eye view of necessary steps for data quality improvement and thus are highly generic. The arguably widest used one is Wang's Total Data Quality Management (TDQM) [9] (see Figure 1). TDQM is suited for continuous monitoring and is open to system evolution - both capabilities important for data quality monitoring in the medical domain. Therefore, we make the assumption that being able to support the TDQM steps is a minimum requirement for any data quality solution to be used by practice managers.

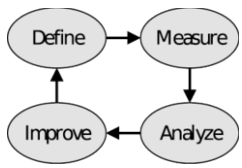


Figure 1 - TDQM cycle

There are many publications specifically considering data quality in healthcare applications, for example health simulations [10], free text analysis of diagnoses [11], or healthcare processes [12]. Concerning data completeness, Miller et al. report about two studies that found "pieces of information perceived as being needed for clinical decisions were missing 13.6% and 81% of the time" [13]. The ACM Journal of Data and Information Quality devoted a special issue to the topic of information quality in healthcare [14]. In 2013, Dixon et al. described the Health Data Stewardship framework and called for software tools to monitor and improve data quality [15]. It follows that supporting high data quality in the medical sector is an ongoing effort.

Interviews

All practice managers reported occurrences of low quality data, with incomplete data being identified as the biggest issue. To be more specific, *population completeness* [1] was deemed the most impactful [16]; a center's practitioners operate on a budget. Once they surpass a certain threshold, their activities are only reimbursed fractionally. This becomes evident at the latest at the end of each fiscal quarter, when practitioners report their activities to a central authority for reimbursement. At this point, however, it is too late for the practice manager to counter this effect. While countersteering would have been possible during the quarter, incomplete data at that time obfuscated arising budgeting problems, leading to loss of revenue.

All practice managers reported that at their centers there already were rules in place about data. Most of these were not checked automatically and had to be evaluated manually. One location reported presence of 119 standardized queries to review some quality constraints, again with the restriction that these queries had to be triggered by hand [17]. This produces significant effort since these queries have to be evaluated frequently. There are some exceptions to this, for example automatic sanity checks performed by patient-data management systems. Still, all practice managers expressed interest in being able to define their own rules for automatic evaluation.

Monitoring Architecture

While our architecture (see Figure 2) is built on top of the MEDITALK packages, it is designed to allow implementation as a standalone application. All user interaction happens through the Interaction package. The Data Service package is responsible for accessing the central database and for storing monitoring results and metadata. Actual data quality measurements are performed by the Monitoring package.

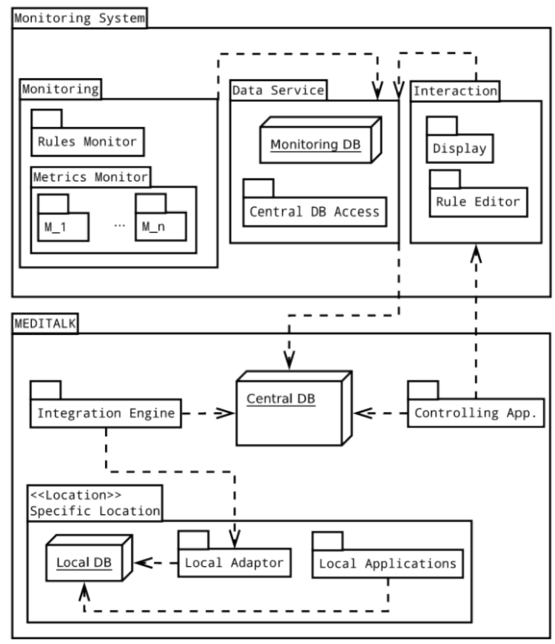


Figure 2 - MEDITALK Integration Environment and Monitoring System

Monitoring Package

Our proposed architecture splits assessment of data quality into two separate concepts: *rules* and *metrics*.

A *rule* is a formulation of a constraint on data, for example R1: ICDCode = 'J00' and Diagnosis != 'Acute nasopharyngitis'[1]. Any tuple in the database for which R1 holds (meaning that the ICD code has an attached diagnosis different from the standard) may be erroneous according to the formulated rule (see [18] for other examples). Another possible form of rules are association rules [19], and conditional functional and inclusion dependencies (see [20] for a survey). Expressing data quality constraints by rules has the advantage that they can be created without intervention by a programmer. Domain experts can formulate rules according to their own requirements.

The **Rules Monitor** package keeps track of the rule base stored in the Monitoring Database, executes rules according to their schedule and in turn stores the results. Rules fall into one of two categories: aforementioned examples are *data quality rules*, and define the fitness of data entries. The second category are *improvement rules*. These are used to trigger events to improve data quality, for example alerting the practice manager to violating data or performing automatic edits[2]. All rules are checked on the available data according to a set schedule. For the MEDITALK project, it is enough to check them once every 24 hours because data arrives in daily intervals. For other contexts, the intervals can be changed according to need. Every rule can also be triggered manually.

[1] This is not necessarily a practical example – we only use R1 to exemplify our system's usage.

[2] Automatic edits bear some challenges like conflicting edits [21]. We assume the practice managers to be aware of these perils.

The **Metrics Monitor** contains all functions that implement metrics, and triggers recalculation according to the needs of the respective metrics. A *metric*, for our purpose, is defined as any measure indicating the quality of a data item or set of data items. In distinction to the mathematical definition of a metric, we denote not only a distance between a data item and its theoretical perfect quality counterpart by the expression "metric", but any indicator which makes a statement about data or data quality. As an example, the number of NULL values within a database table may indicate the table's completeness (depending on NULL semantics [22]). We use the term metric interchangeably as the indicator delivered or the mechanism to calculate the indicator.

We distinguish between metrics and rules for the following reasons: some desirable functions, for instance statistical models (see [23] for an example), may be too complex for easy representation by rules. Also, some rules may prove to be of interest to more than one center. In the case of large sets of rules, creating these rules at all interested centers would be redundant. Expending onetime effort encoding or generating the set in a metric, then integrating the resulting module at each center would lower overall effort. So if for these reasons additional programming is warranted, an additional metric is created. This will usually be done by a programmer in response to a new requirement by a domain expert.

Data Service Package

All results of calculations within the Monitoring package are stored in the **Monitoring Database** to encapsulate all information necessary to assess gathered indicators. A monitoring result is a statement about data, valid at a certain point in time. Each result is identified by;

- the originating rule's / metric's identifier,
- the date and time the result was obtained, and
- all involved tuples and attributes.

Apart from this identifying data, each issue has a payload according to the specific rule / metric, for example an indicator calculated by a metric. The payload may be empty if all necessary information is already present in the result's key, but may also be arbitrarily complex, for example delivering an XML data structure. Usually, a rule's payload will be the percentage of applicable tuples it holds on, and the identifiers of violating tuples. The payloads of metrics may be as simple as a single indicator (e.g. "expected number of patients tomorrow: 32"), possibly including a confidence interval for the value. They may also deliver structures of any number of values and of any depth (e.g. a decision tree derived from patient data).

The **Central Database Access** package serves as a mapper between the Central Database and the data requirements of the Monitoring System. This decouples the Central Database and all other packages that need access to it, preventing changes at the Central Database from propagating to other parts of the system.

Interaction Package

Through the **Rule Editor**, practice managers can define their own data quality rules. This solution has several advantages in comparison to the standardized queries used to date:

- Rules can still be triggered manually, but can also be checked automatically and periodically (at set execution intervals).
- Rules can not only be defined on data, but also on metadata, meaning that rules can be used to alert the practice manager when a metric (or even another rule) changes.
- Drag-and-drop generation of rule conditions may lower initial training effort compared to explicitly spelling out the queries.

The prototypical rule editor (see Figure 3) developed in our group [24] can be used to create Boolean rules, for example to check whether a value matches a regular expression. Users can write their own regular expressions, or choose from standard checks like string length or number format. The editor is able to recreate the standardized queries, and can be extended to allow (conditional) functional dependencies and other kinds of rules not yet implemented.

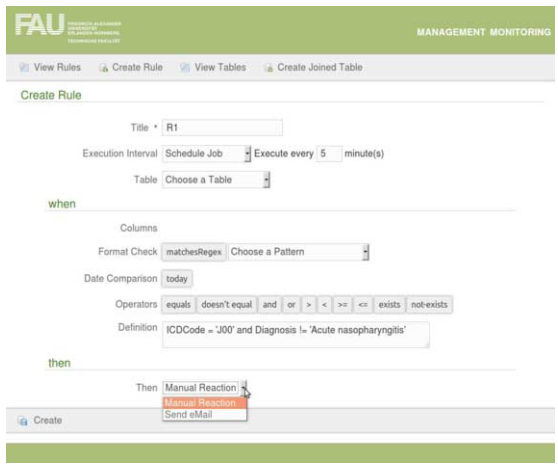


Figure 3 - Rule creation

The **Display** package is responsible for visualizing the state of the Monitoring System. This includes all existing rules and metrics as well as their results on the currently available data. It also enables exploratory analysis of the central database to get an overview over schema and extension. All subpackages of Interaction have their own user interfaces, which can be used on their own or be implemented as plugins for the controlling application in the MEDITALK package.

Discussion

To evaluate whether our architecture is capable of continuous data quality monitoring, we show that each step of TDQM is supported by one or more packages of our architecture, as stated in Figure 4.

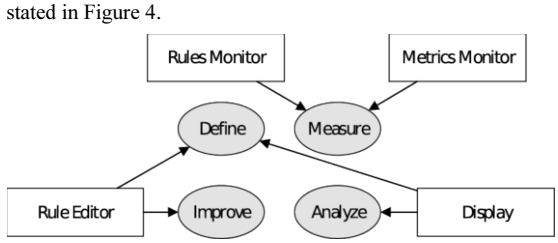


Figure 4 - Supporting TDQM

Define

"The definition component of TDQM cycle identifies important data quality dimensions (...) and the corresponding data quality requirements." [7, p.5, "The TDQM Cycle"]

Defining fitness of data is closely tied to an organization's business goals. As such, some of the effort expended in this step is outside of the scope of the Monitoring System. The

Display package assists in this effort by delivering an overview over available data and the schema of the central database. The formal definition of data quality requirements however is supported by the **Rule Editor** in full.

Example A practice manager (PM in the following) discovers that at some locations, ICD code and diagnosis text are both entered manually. The PM decides that mismatches between ICD code and diagnosis should be avoided, and formulates rules to check for these mismatches. One of these rules is R1: ICDCode = 'J00' and Diagnosis != 'Acute nasopharyngitis'. A catalog of like constraints to check the quality requirement "ICD code and attached diagnosis may not mismatch" is defined.

Measure

"The measurement component produces data quality metrics." [7, p.5, "The TDQM Cycle"]

The measure step is performed by the **Rules** and **Metrics Monitoring** packages. Both calculate indicators for data quality and store them in the Monitoring Database.

Example A new practice joins the center. Once its data becomes available in the central database, Rules Monitoring checks whether all applicable rules hold. Assuming the practice uses the diagnosis "common cold" instead of "Acute nasopharyngitis" in 5 cases with ICDCode = 'J00', all of these fulfill R1. A result is placed in the Monitoring Database stating that 5 tuples (the ones coming from the new practice) fulfill R1 and therefore are of low quality.

Analyze

"The analysis component identifies root causes for data quality problems and calculates the impacts of poor quality information." [7, p.5, "The TDQM Cycle"]

The Analyze step is supported by the interaction of the **Monitoring Database** and the **Display** package. The Display package shows all Monitoring results. It allows drilling down into their information, showing affected tuples and attributes, their provenance, and involved rules and metrics.

Example The result R1 match ratio: 0.13% is shown in the Display package (see Figure 5). This means that out of all diagnosis entries on the central database, 0.13% violate the data quality requirement defined by R1. The PM checks the definition text of the rule as well as which tuples are involved and their provenance. In this way, the new practice is identified as the origin of the problem.

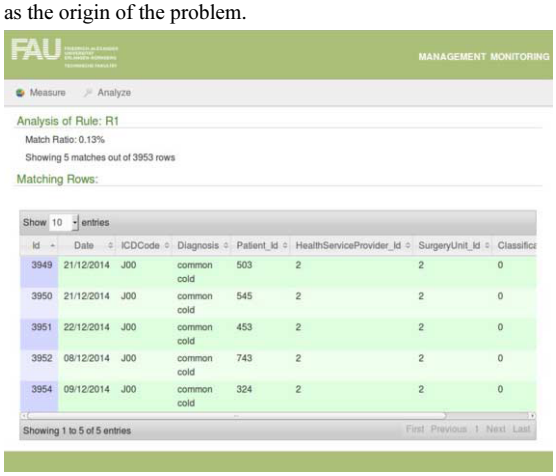


Figure 5 - Analyzing a data quality rule

Improve

"(...) the improvement component provides techniques for improving data quality." [7, p.5, "The TDQM Cycle"]

Again, some of the actions performed during this step are outside of the scope of the Monitoring System (e.g. organizational changes or in-service training of staff). However, the **Rule Editor** can be used to establish improvement rules. An example of such a rule is sending an email to a location that has not entered data by a certain deadline. Data cleaning rules such as automatic edits can also be formulated.

Example Judging the low data quality to be a one-time problem, the PM instructs the new practice to revise the offending tuples. Should the problem occur again, an improvement rule can be created that on detection of a violating tuple either notifies the person responsible or automatically changes the values to conform.

Conclusion

We described an architecture supporting continuous data quality monitoring for medical centers. Lenz identified several important factors in making software sustainable [25]. Our architecture satisfies two of these by design. *Separation of Concerns* is granted by grouping functionality into fitting packages, separating rules, metrics, data access, and user interaction into their own packages. This minimizes the probability that change in one package requires redesign of others. *Deferred Design* is satisfied by the presence of the rule editor and rule monitoring: The set of rules is not hardcoded into the system, but can be built, corrected, and extended at runtime by domain experts as demand requires. This ensures evolvability of the rule system with changing requirements. Two additional factors, *Loose Coupling* and *Service Oriented Architecture*, are implementation specific.

While our application example is an integration scenario, this is not imperative. Since the architecture only requires access to a single central database and not the sources, it can just as easily be applied to databases that are not involved with integration, for example hospital information systems. This design does not diminish the potential to make use of information that arises in the ETL process: Any data and metadata that is stored in the connected database is fair game for monitoring.

By enabling continuous monitoring, system evolution and deferred design by domain experts, we also support frameworks like Health Data Stewardship (HDS) [15]. The connection to health outcomes that HDS demands is implemented by the system's ability to monitor any kind of data. Since outcomes may also be stored in the central database, any information about those can be used by rules and metrics as well as in the display component.

Several parts of the architecture are already implemented. We are currently developing methods for measuring population completeness. Access to the central database is handled directly through the accessing components and not through a common mapper as of now. Monitoring DB and Display are not yet complete, but are already capable of storing respectively presenting the results of rules. Rule Editor and Rule Monitoring are fully functional. Once complete, we will evaluate the effectiveness of our solution through user studies.

Acknowledgements

Parts of this work are supported by the German Federal Ministry of Education and Research (BMBF), grant 13EX1013D.

References

[1] Batini C, Scannapieco M. Data Quality: Concepts, Methodologies and Techniques. Springer; 2006.

[2] Endler G. Data quality and integration in collaborative environments. In: Proceedings of the SIGMOD/PODS 2012 PhD Symposium. New York, NY, USA. p. 21–26.

[3] Parnas DL. Software aging. In: Proceedings of the ICSE '94. Los Alamitos, CA, USA: IEEE Computer Society Press; 1994. p. 279–287.

[4] Lenz R, Kuhn KA. Towards a continuous evolution and adaptation of information systems in healthcare. International Journal of Medical Informatics. 2004;73(1):75 – 89.

[5] Endler G, Langer M, Purucker J, Lenz R. An Evolutionary Approach to IT Support for Medical Supply Centers. In: Proceedings der 41. Jahrestagung der Gesellschaft für Informatik e.V. (GI); 2011.

[6] Chaudhuri S, Dayal U. An overview of data warehousing and OLAP technology. SIGMOD Rec. 1997 March;26:65–74.

[7] Wang RY, Ziad M, Lee YW. Data Quality. Springer US; 2002

[8] Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM Computing Surveys. 2009 July;41:16:1–16:52.

[9] Wang RY. A product perspective on total data quality management. Communications of the ACM. 1998 February;41:58– 65.

[10] Baumgärtel P, Lenz R. Towards Data and Data Quality Management for Large Scale Healthcare Simulations. In: Proceedings of the International Conference on Health Informatics. 2012. p. 275–280.

[11] Lauria EJM, March AD. Combining bayesian text classification and shrinkage to automate healthcare coding: a data quality analysis. J Data Inf Qual. 2011;2(3):13:1–13:22.

[12] Lenz R, Reichert M. IT support for healthcare processes - premises, challenges, perspectives. Data Knowl Eng. 2007;61(1):39-58.

[13] Miller DW, Yeast JD, Evans RL. Missing Prenatal Records at a Birth Center: A Communication Problem Quantified. In: AMIA Annu Symp Proc. American Medical Informatics Association; 2005.

[14] Madnick SE, Lee YW, editors. Special Issue on Information Quality: The Challenges and Opportunities in Healthcare Systems and Services. vol. 4 Issue 1 of ACM JDIQ. New York, NY, USA: ACM; 2012.

[15] Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. Proceedings of the 14th World Congress on Medical and Health Informatics. 2013;192:884–888.

[16] Endler G, Baumgärtel P, Lenz R. Pay-as-you-go data quality improvement for medical centers. In: Proceedings of the eHealth 2013; 2013.

[17] Gorupec M. Entwicklung eines Frontends für Regelsysteme zur Datenqualitätsverbesserung medizinischer Informationssysteme. Master Thesis, FAU; 2014.

[18] Rakovac I, Maharjan B, Stein C, Loyola E. Program for validation of aggregated hospital discharge data. Proceedings of the 14th World Congress on Medical and Health Informatics. 2013;192:1155.

[19] Agrawal R, Imielinski T, Swami A. Mining Association Rules Between Sets of Items in Large Databases. In: Proceedings of the 1993 ACM SIGMOD '93. New York, NY, USA: ACM; 1993. p. 207– 216.

[20] Fan W, Geerts F, Jia X, Kementsietsidis A. Conditional functional dependencies for capturing data inconsistencies. ACM Trans Database Syst. 2008 Jun;33(2):6:1–6:48.

[21] Fellegi IP, Holt D. A systematic approach to automatic edit and imputation. J Am Stat Assoc. 1976;71(353):17–35.

[22] Zaniolo C. Database relations with null values. In: Proceedings of the 1st ACM SIGACT-SIGMOD symposium on Principles of database systems. PODS '82. New York, NY, USA: ACM; 1982. p. 27–33.

[23] Huang Y, Hanauer DA. Patient No-Show Predictive Model Development using Multiple Data Sources for an Effective Overbooking Approach. Appl Clin Inform. 2014;5(3):836–860.

[24] Gorupec M, Endler G. ruleDQ: Ein Regelsystem zur Datenqualitätsverbesserung medizinischer Informationssysteme. In: GI Lecture Notes in Informatics (LNI) Seminars 13 / Informatiktage 2014; 2014. p. 37–40.

[25] Lenz R. Information Systems in Healthcare - State and Steps towards Sustainability. IMIA Yearbook 2009. 2009;1:63– 70.

Address for Correspondence

Gregor Endler

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Computer Science 6 (Data Management)

Martensstraße 3, 91058 Erlangen, Germany

gregor.endler@fau.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-857

A Conceptual Framework for Decision-making Support in Uncertainty- and Risk-based Diagnosis of Rare Clinical Cases by Specialist Physicians

Adriano A. Santos[a], J. Antão B. Moura[b], Joseana Macêdo Fechine Régis de Araújo[b]

[a] *Graduate Program in Computer Science from the Federal University of Campina Grande - UFCG*
Centro de Ensino Superior e Desenvolvimento – CESED Campina Grande - PB, Brazil
[b] *Systems and Computing Department - Federal University of Campina Grande – DSC/UFCG, Brazil*

Abstract

Mitigating uncertainty and risks faced by specialist physicians in analysis of rare clinical cases is something desired by anyone who needs health services. The number of clinical cases never seen by these experts, with little documentation, may introduce errors in decision-making. Such errors negatively affect well-being of patients, increase procedure costs, rework, health insurance premiums, and impair the reputation of specialists and medical systems involved. In this context, IT and Clinical Decision Support Systems (CDSS) play a fundamental role, supporting decision-making process, making it more efficient and effective, reducing a number of avoidable medical errors and enhancing quality of treatment given to patients. An investigation has been initiated to look into characteristics and solution requirements of this problem, model it, propose a general solution in terms of a conceptual risk-based, automated framework to support rare-case medical diagnostics and validate it by means of case studies. A preliminary validation study of the proposed framework has been carried out by interviews conducted with experts who are practicing professionals, academics, and researchers in health care. This paper summarizes the investigation and its positive results. These results motivate continuation of research towards development of the conceptual framework and of a software tool that implements the proposed model.

Keywords:

Conceptual framework; Risk; Uncertainty; Clinical Decision Support Systems.

Introduction

Clinical Decision Support Systems (CDSS) provide medical professionals with knowledge at appropriate time, and in appropriate form. In general, they ease decision-making process, making it efficient and effective, helping to reduce the number of avoidable medical errors and enhancing quality of treatments provided to patients [1].

Having this in mind, much research has been developed in order to provide both specific solutions (Specialized Systems) [1], [6], [7] and general ones (conceptual frameworks) [2], [3], [4], [11] that can help Specialist Physicians in reducing the number of medical errors; reducing duration of appointments; managing knowledge or even creating new clinical procedures which may become a standard for the entire medical team. Nevertheless, there is a category of clinical cases which deserve more caution: the rare cases.

According to Loriggio [8], when a problem presents difficulties associated to the unavailability of solution or demands too many resources (money, staff, time, information about environmental elements, etc.) and, therefore, too much effort for solving or diagnosing its causes, such a problem is said to be complex or rare. We can infer that these problems are directly connected to lack of: knowledge about the rare case; reliable and available sources; sharing and exchange of knowledge and learned cases; more precise analyses; and evidence raising.

Day after day, specialist physicians carry out complex health analyses and must make decisions which might be fatal, in case they were erroneous. Very often, rare cases are found and methods that allow reduction of uncertainty about procedures must be executed in order to settle risks associated with the decision. Urgent clinician-clinician communications require routes of contact that are fast and dependable and allow for exchange of complex information [10].

In an attempt to contribute to the solution of rare clinical cases, this work tries to answer the following questions: how can one mitigate risks and uncertainty that challenge specialist physicians when analyzing an unfamiliar clinical case? Which technologies and/or techniques could be harnessed to create a conceptual framework to give support to decision-making process in a clinical environment?

This paper proposes a conceptual framework for helping in the analysis and decision-making process of rare clinical cases with basis on the association of medical evidence analysis techniques, knowledge management, and collective intelligence in order to mitigate the risks and uncertainty faced by specialist physicians. Another important characteristic of the proposed framework is as a controlled teaching and learning tool with the possibility of formalizing a process for aiding training of specialist physicians,.

In fact, we anticipate that the most likely use of our model will be in the education, to help train health professionals. The educator can set up controlled cases and submit them to physicians being trained, so that they can give their opinions, in a simulated environment. Thus building knowledge by means of interactions among the students. At the end, the educator can evaluate the interactions and even suggest more sources or procedures in order to consolidate or generate more significant knowledge.

Hypotheses

We tested the following hypotheses in this research:

H_a: The proposed framework will be of help in the decision-making process in rare clinical cases.

H_b: The proposed model can be used as a tool for knowledge management and for professional training of specialists.

H_c: Decision makers are satisfied with the proposed framework and can identify effective contributions to this area.

Related Work

The first studies that we can refer to, in the context of this research, were conducted by Dobrow et al. in 2004 [2], first version, and Dobrow et al. in 2006 [3], an evolution of the previous version. The authors´ decision-making approach is based on evidence techniques in public health. They defend that evidence and context are integral and fundamental components for evidence-based decision process. The authors present a conceptual framework which aims to analyze the impact of use of context in constitution of evidence and how evidence is used in medicine. They make use of an "evidence source" or, technically, a knowledge base, and consider the role of context in the introduction, interpretation, and application of evidence to support the decision-making process.

The studies conducted by Dobrow et al. in 2004 [2] and Dobrow et al. in 2006 [3] are related to this research because they deal with a conceptual framework and use the evidence concept. However, they are limited to focusing on how context impacts on what constitutes evidence and how that evidence is utilised. They do not use other knowledge sources or processes to improve the decision-making process.

In work by Ferreira et al. in 2007 [4], a framework for the decision process in emergency cases is presented. The authors use database management systems (DBMS), a specialist's base, and a geographic information system (GIS); to plan tasks which will help reduce the time necessary for strategic planning in response to emergency cases. The framework also features an auxiliary module for simulation of cases. The work by Ferreira et al. [4] is similar to our proposal due to the use of a specialist database and to the purpose of reducing response time of clinical cases. However, the knowledge sharing process, the treatment of uncertainty and the management of learned cases presented in our proposal were not covered by the authors' work.

Lopes et al. 2009 [7], proposed a decision support conceptual framework based on evidence, context, and cases called EcoCADe (Evidence, Context, and Cases to Support Decision). The authors proposed a hybrid solution in GIS and CBR (Case Based Reasoning), allowing specialists to model evidence of PBE (Evidence-Based Practice) domain, both using problem context. The authors proposed a solution that does not depend on a specific domain, which allows its adaptation to several domains. A methodological approach to support the decision-making process involved RBC techniques and/or PBE procedures. The authors validated the model both in legal and medical contexts. It is a robust and adaptive solution which uses concepts that are part of the basis of our work. Nevertheless, risk and uncertainty factors, collective knowledge process as support to decision-making process, which are present in our work, were not covered by those authors' work.

Kastner and Straus 2012 [6], proposed a conceptual framework used in the development of a support tool for clinical decision of osteoporosis cases. The authors exposed security problems in the use of information systems in the process of supporting decision-making in clinical cases. They developed a solution that uses Knowledge-to-Action cycle and stages of the Medical Research Council framework. That work is of interest to our research because it raises a set of issues and possible problems in the use of information systems as support tools for the decision-making process in medical systems. The authors created an expert system with the objective of supporting diagnosis of osteoporosis problems. We do not propose development of an expert system, but a domain-independent framework, such as EcoCaDe. Even the development of an IS solution for future validation, could be made flexible and adaptable to the desired domain.

Minutolo and collaegues 2012 [9], proposed a knowledge editing system based on standards to direct and help in creation and formalization of clinical recommendations, to be used in knowledge-based Decision-Support Systems. That system was conceived to offer a set of standards that ease insertion and editing of clinical recommendations, using the combination of knowledge representation techniques to instantiate standards in knowledge bases, using rules built by using ontologies. That research is related to our work because it involves a hybrid solution that also enables management of knowledge generated by the environment. However, as in the work of Kastner and Straus [6], it is a specialized software.

Table 1 summarizes a comparison of this paper and related works reviewed here.

Table 1– Comparison of reladed work.

Work	Independent Model and Extensible Domain	Collective Intelligence	Treatment of Risks and Uncertainties	Knowledge management	Training
[2,3]	X			X	
[4]			X	X	
[7]	X			X	
[6]	X		X		
[9]	X		X	X	
Ours	X	X	X	X	X

The Proposed Conceptual Framework

Our framework comprises five macro stages, presented in Figure 1, namely: a) Anamnesis Process; b) Knowledge Management; c) Decision Strategy; d) Application of Procedures; and e) Learned Lessons.

Each stage presented in the framework comprises a set of pooled techniques which receives, sequentially, the outputs of previous stages. The specialist physician in charge of the analysis of the case is responsible for the Anamnesis, Knowledge Management, and Application of Procedures stages. The Decision Strategy stage makes use of Delphi method in an attempt to achieve convergence of the opinions of experts in the decision-making process. The end of the process is the generation of a new case, which is stored on the Learned Lessons database.

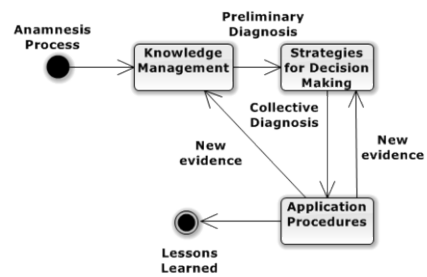


Figure 1– The conceptual framework

In the Anamnesis Process stage, the specialist physician interviews the patient in order to gain as much information as possible about her condition. This stage comprises three steps, namely: a) Questions (Open, Focuses – or Semi-open, Closed, Directed, and Composite); b) Collecting (Answers); and c) Observed evidences (previous analyses based on the survey carried out). The output of this process is a clinical evidence list, which will be used as input by the Knowledge Management stage.

In the Knowledge Management stage, the specialist physician checks the evidence collected in the Evidence step in two knowledge sources, attempting to find consolidated evidence that may serve as basis for a preliminary diagnosis.

The first knowledge source used is based on reliable external sources, and uses an evidence search engine. The second knowledge source is a knowledge base formed by the set of lessons learned with the experiences of the solved cases and the experience of the specialist physicians.

The output of the Knowledge Management stage is a Preliminary Diagnosis of the specialist physician. This information is sent to the third stage of the process, called Decision-Making Strategies.

In the Decision-Making Strategies stage, preliminary diagnosis is evaluated by a set of specialist physicians, characterizing the Collective Intelligence process. The formalization of this stage uses Delphi method [5], so that the expert does not make biased decisions for reasons such as marriage and friendship with other experts, and especially because it will be possible to share opinions among themselves to the point of reaching a consensus.

The decision-making task of a specialist physician faced with a rare situation is highly risky, as said in the introduction of this work. However, we believe that, joining the Delphi method and the Collective Intelligence process, as proposed in this stage, with the analysis of the contributions of other specialist physicians will mitigate the risks, because these specialist physicians will contribute their practical experiences, since they have knowledge about the studied domain.

The contribution of experts is possible due to the controlled feedback that defines regular communication between participants in the discussions of previous rounds summaries, reducing noise in communication and avoiding the diversion of the main discussion of the problem. During the rounds of discussions, experts receive information and comments of others, and may change or provide their thoughts with more appropriate arguments [5].

The output of the Decision-Making Strategies stage is a collective diagnosis, which will be used by the specialist physician, who started the process, in the Application of Procedures stage.

The Application of Procedures stage, is actually where the specialist physician applies the clinical procedure to the patient. The result of this stage may generate other evidence which, according to the specialists' decision, can be fed back to the Knowledge Management or the Decision-Making Strategies stages; in case the procedure proposed by the collective diagnosis is not successful (e.g: medications and drug allergic reactions). However, if the procedure is successful, the whole case becomes a Learned Case and it is stored on the Learned Lessons base, with the possibility of being used as a source of information for new cases.

Clinical Decision Support Systems for Analysis of Unfamiliar Clinical Cases (CDSS-AUCC)

Implementation of the Clinical Decision Support for analysis of unknown clinical cases (CDSS-AUCC) tool (Figure 2), including all modules in the proposed framework, is ongoing. CDSS-AUCC will be fully Web compatible and make use of a Business Process Model and Notation (BPMN) service to business rules. The system will be evaluated in a case study. Later, it will be used as a decision-making support tool in a school clinic; and also as an environment aid in the education and the training of health experts at a university in the state of Paraíba, Brazil.

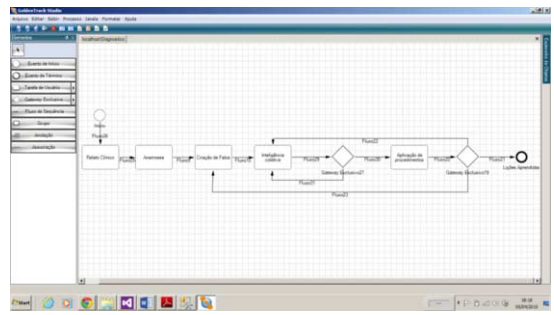


Figure 2–Clinical Decision Support for analysis of unknown clinical cases (CDSS-AUCC)

As of writing this paper, a first version for each of the modules that represent the processes of: a) Anamnesis Process (Figure 3); b) Knowledge Management (Figure 4); c) Decision Strategy (Figure 5); and d) Application of Procedures (Figure 6) are already implmented. Leaving only the Learned Lessons. Development is ongoing to produce evolutionary – in terms of functionality and performance – versions for each module.

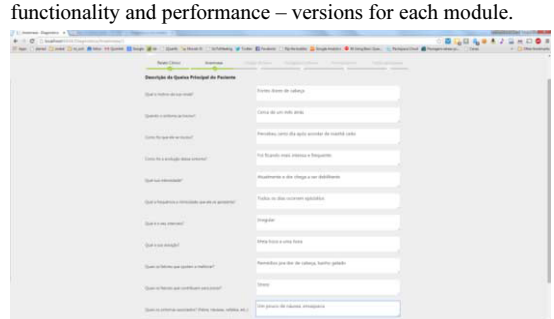


Figure 3– Anamnesis Process

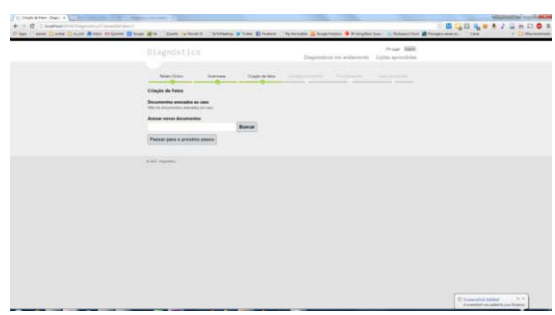


Figure 4– Decision Strategy

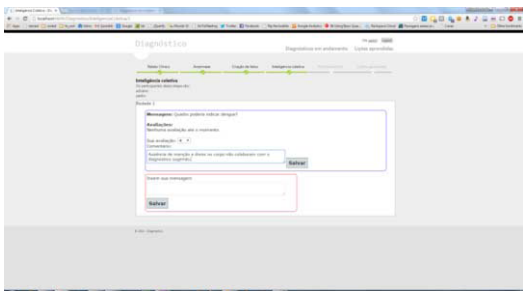


Figure 5– Knowledge Management

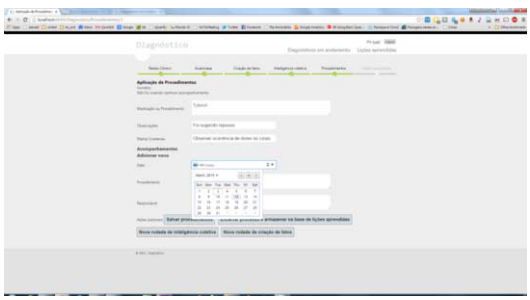


Figure 6– Application of Procedures

Methods

This research is classified as an applied and qualitative research. It aims to get the opinion of a group of health experts (target audience of the research) on the feasibility and the applicability of the proposed framework.

The results of the literature review helped to develop questions for the semi-structured interview protocol presented. Then, data were collected through a survey and interviews for experts' perspectives and recommendations on the dissemination of computer decision support software and the conceptual framework.

Interviews were conducted with 20 experts who work in Brazil's northeastern state of Paraíba as health professionals, academics, and researchers. Practitioners were, preferably active in private or public hospitals. Among the selected professionals were eight cardiologists, two geneticists, four mental nursing specialists, two physiotherapists, two specialists in gastroenterology, and two ophthalmologists. Three specialists have private clinics and four coordinate undergraduate courses in medicine in their respective educational institutions. Four of the experts coordinate scientific research projects.

We wanted to know their opinions and formalize these opinions into a document which would guide us through the rest of the research. The questions presented to them were:

A. Can we believe the proposed framework will be of help in the decision-making process in the diagnosis of rare clinical cases?
B. Can the proposed model be used as a tool for knowledge management and for professional training of specialists?
C. As a decision maker, are you satisfied with the proposed model and could you identify any effective contribution to this area?
D. Would you use an IT tool based on this model to support your decisions?
E. Do you consider the model useful and would you invest (time, specification effort, etc.) in its evolution?

We and the interviewed experts used the Think Aloud Protocol [12]. In addition to offering comments and suggestions, the experts answered the questionnaire with questions A) to E) above. Possible questionnaire responses were in the form of a 4-level Likert scale: Strongly disagree = 0; Disagree = 1; Agree = 3; and, Strongly agree = 4.

Experimental Unit

As an experimental unit, we selected a group of experts in health who work in private and public clinics in northeastern Brazil. All participants are professors of medical schools at public and private universities, with at least a master degree (eight have doctoral degrees), and minimum professional experience of five years. Furthermore, 70% of them also function as researchers.

Experimental Design

Twenty interviews were performed as described. Each interview lasted an average of two hours. Interviews of experts were conducted in two steps: in the first step, after the presentation of the framework and the system, the interviewees reported their opinions orally and subsequently questioned about the system usability. For this step, we used the principles of the Think Aloud protocol. In the second step, the experts answered the questionnaire. This questionnaire was created to analyze the prevalence of opinions of experts over the research objectives and also give answers about the hypotheses for this research.

Results

Statistics

We used an ordinal scale on the answers to questions A, B, and C; performed median tests (Wilcoxon test) with the degree of reliability of 95%, and the ratio test. The results are shown in Table 2.

Table 2– Statistical tests results

Question	V	p-value	Delta	Conf. Interv.
A	91	0.0001	0.209037	0.44 - 0.85
B	78	0.0003	0.214703	0.38 - 0.81
C	65	0.011	0.219130	0.28 - 0.71

All the experts said they would use an IT tool based on the model and confirmed interest in contributing to the development of research.

Graphics

The results of the experiment for questions are given in figure 7.

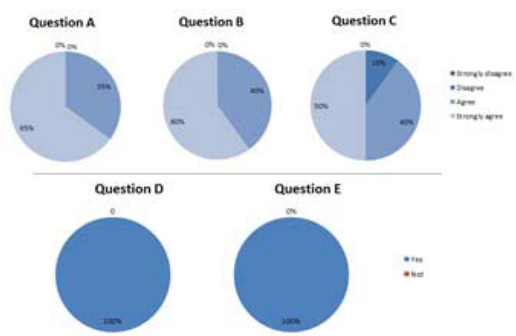


Figure 7– Results of the experiment

Discussion

The main aim of this research is to analyze whether the proposed framework has an effective contribution in the decision-making process in unfamiliar clinical cases. Therefore, hypothesis H_a, H_b, and H_c were raised and submitted to representative experts, so they could give their appraisal.

For preliminary validation of the usefulness of the proposed framwork, we interviewed twenty health professionals who function as professors, researchers, and practitioners in medical area. They went through two stages of analysis (Think Aloud and Face Validity) and answered a questionnaire to a) assess whether they *thought* the proposed framework was valid for their decision-making process, b) whether it would be useful in the teaching process and c) continuing training of new specialists; and whether there would be scientific contribution to the field of rare cases diagnosis.

According to the results found in the processing of statistical tests and the descriptive data analysis, one may conclude that, in the opinion of experts, the conceptual framework was approved as a contribution to the medical area and it will bring benefits to the educational process, training of new specialists, and to the professional field in general.

The development of a solution that implements the proposed framework was also considered promising since all the involved specialists offered to participate in the development and evolution process of the research.

The group of experts selected for this research, came from different backgrounds, interests and specialties which made for a rich and heterogeneous sample in perceptions. They put forward their real needs and directly contributed to the evolution and development of the research.

Thus, according to all the points raised, the null hypotheses is rejected ($p < 0,05$). Therefore, stimulating continuation and development of the research.

Conclusion and Outlook

This research aims to explore the opinions of experts in the health area on a conceptual framework to support decision making, mitigating risks and uncertainties faced by experts in unfamiliar clinical cases. The conceptual framework proposed in this paper is based on five stages: Anamnesis Process, Knowledge Management, Decision Strategy, Application of Procedures and Lessons Learned.

For preliminary validation of the usefulness of the proposed framwork, we interviewed twenty health professionals who function as professors, researchers and practitioners in medical area. They went through two stages of analysis (Think Aloud and Face Validity), and answered a questionnaire to assess whether they *thought* the proposed framework was valid for their decision-making process; whether it would be useful in the teaching process and continuing training of new specialists; and whether there would be scientific contribution to the field of rare cases diagnosis. Their positive answers motivate continuation of the research; and development of the conceptual framework which will lead to availability of a software tool that implements the proposed model.

As for future work, we will concentrate on the completion of the development of the CDSS-AUCC tool – a software tool that implements the proposed framework – and on the study of cases with more clinical experts for further validation of the model.

References

[1] Berner ES. Clinical decision support systems: theory and practice. 2nd ed. New York: Springer; 2007.

[2] Dobrow MJ, Goel V, Upshur REG. Evidence-based health policy: context and utilization. Soc Sci Med 2004; 58(1): 207-217.

[3] Dobrow MJ, Goel V, Lemieux-Charles L, Black NA.The impact of context on evidence utilization: a framework for expert groups developing health policy recommendations. Soc Sci Med 2006; 63(7): 1811-1824.

[4] Ferreira F, Dantas A, Seville E. Conceptual dynamic response recovery tool for emergency events. In: The International Forum on Engineering Decision Making (IFED) Third Forum. 12-15 Dec. 2007; Port Stephens, Australia. University of Canterbury: Civil and Natural Resources Engineering; 2007.

[5] Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. Practical Assessment, Research & Evaluation 2007; 12(10): 1-8. Available in: <http://pareonline.net/pdf/v12n10.pdf>. Accessed in 2014

[6] Kastner M, Straus SE. Application of the Knowledge-to-Action and Medical Research Council frameworks in the development of an osteoporosis clinical decision support tool. J Clin Epidemiol 2012; 65(11):1163-1170.

[7] Lopes EC, Schiel U, Santos Jr., Gilson P. A decision support methodology for the control of alternative penalties - a case-based reasoning approach. International Conference on Information, Process, and Knowledge Management (eKNOW 2009). IEEE Computer Society. 2009

[8] Loriggio A. De onde vêm os problemas. São Paulo, Brasil: Negócio Editora. 2002

[9] Minutolo A, Esposito M, De Pietro G. A pattern-based knowledge editing system for building clinical decision support systems. Knowledge-Based Systems 2012; 35: 120-131.

[10] Nguyen C, McElroy LM, Abecassis MM, Holl JL, Ladner DP. The use of technology for urgent clinician to clinician communications: a systematic review of the literature. Int J Med Inform 2015; 84(2): 101-110.

[11] Sanchez E Toro C, Artetxe A, Grana M, Sanin C, Szczerbicki E, Carrasco E, Guijarro F. Bridging challenges of clinical decision support systems with a semantic approach. A case study on breast cancer. Pattern Recognition Letters 2013; 34(14): 1758-1768.

[12] van Someren MW, Barnard YF, Sandberg JAC. The think aloud method: a practical guide to modeling cognitive processes. San Diego, CA: Academic Press Ltd. 1994

Acknowledgment

This work is being supported in part by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES).

Address for correspondence

Adriano Araújo Santos – adriano@copin.ufcg.edu.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-862

A Hybrid Approach Using Case-Based Reasoning and Rule-Based Reasoning to Support Cancer Diagnosis: A Pilot Study

Renata M. Saraiva[a], João Bezerra[c], Mirko Perkusich[c], Hyggo Almeida[a], Clauirton Siebra[b]

[a] *Department of Computing and Systems, Federal University of Campina Grande, Campina Grande, PB, Brazil*
[b] *Department of Informatics, Federal University of Paraíba, João Pessoa, PB, Brazil*
[c] *Federal Institute of Paraiba, Monteiro, PB, Brazil*

Abstract

Recently there has been an increasing interest in applying information technology to support the diagnosis of diseases such as cancer. In this paper, we present a hybrid approach using case-based reasoning (CBR) and rule-based reasoning (RBR) to support cancer diagnosis. We used symptoms, signs, and personal information from patients as inputs to our model. To form specialized diagnoses, we used rules to define the input factors' importance according to the patient's characteristics. The model's output presents the probability of the patient having a type of cancer. To carry out this research, we had the approval of the ethics committee at Napoleão Laureano Hospital, in João Pessoa, Brazil. To define our model's cases, we collected real patient data at Napoleão Laureano Hospital. To define our model's rules and weights, we researched specialized literature and interviewed health professional. To validate our model, we used K-fold cross validation with the data collected at Napoleão Laureano Hospital. The results showed that our approach is an effective CBR system to diagnose cancer.

Keywords:

Medical Decision Support Systems; Case Based Reasoning; Rules Based Reasoning.

Introduction

According to the World Health Organization (WHO), cancers figure among the leading causes of morbidity and mortality worldwide[1]. Researchers from the WHO and the International Agency for Research on Cancer (IARC) claim that in 2012, there were 14.1 million new cancer cases and a total of 8.2 million deaths from cancer worldwide. It is expected that the number of cancer patients will continue to rise by about 70% over the next 2 decades[1]. Early diagnosis of cancer is a big challenge because it is a disease with multiple locations and clinicopathological aspects while having no pathognomonic (i.e., specific to each disease) signs or symptoms[2]. Therefore, it can be detected in various stages of histopathological and clinic evolution. Healthy lifestyle and early diagnosis of this disease can reduce its mortality rate, according to the International Union of Cancer Control (UICC)[3].

Many researchers have applied Artificial Intelligence (AI) techniques to create health-related systems or models, such as the diagnosis or classification of diseases [1-4]. To diagnose cancer, researchers applied different computational techniques. Salem and El Bagoury [5] proposed a hybrid case-based adaptation model, that combines transformational and hierarchical adaptation techniques with artificial neural networks and certainty factors for the diagnosis of thyroid cancer. Zubi and Saad [6] combined data mining techniques with neural networks for the early diagnosis of lung cancer. For the diagnosis of breast cancer, Keles, Keles and Yavuz [7] used neuro-fuzzy rules while Sharaf-elDeen et al. [8] used a hybrid approach that combined case-based reasoning (CBR) with rule-based reasoning (RBR).

In this paper, we present a hybrid approach using CBR and RBR to assist healthcare professionals in the early diagnosis of patients with cancer. We use CBR and RBR because rules and cases are complementary [9]. Instead of using RBR as an alternative solution to CBR, as Sharaf-elDeen et al. [8] did, we use it to improve the probability of CBR to converge to the best solution. The proposed model may act both as a decision support system for less experienced clinicians and also as a second opinion for experts.

To carry out this research, we received the approval of the ethics committee at Napoleão Laureano Hospital. To define our model's cases, we collected real patient data at Napoleão Laureano Hospital, which is a reference for oncology in Brazil. We represented the case with patients' personal information, signs (i.e., objective findings that can be described by a health-care provider), symptoms (i.e., subjective complaints reported by patients), and their diagnoses. To define our model's rules and weights, we researched specialized literature [10] and interviewed a general practitioner.

For the purpose of this research, we collected data from patients with gastrointestinal cancer. More specifically, patients with the following gastric neoplasms: anal, colorectal, esophagus, and stomach. To validate our model, we developed a prototype and used the K-fold cross validation method. The final results show that our approach has increased the accuracy of the diagnosis by 22.92% when compared to using only CBR.

Background

Case-based reasoning

Case-Based Reasoning is a paradigm for solving problems that is fundamentally different from other major AI approaches. Instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to use specific knowledge of previous experiments from concrete problems (cases) [11]. In CBR, a new problem is solved by reusing the solution of a previous

[1] http://www.who.int/mediacentre/factsheets/fs297/en/
[2] http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/estomago/diagnostico_profissional
[3] http://www.uicc.org/national-cancer-leadership-congress-2014

similar problem. A second important difference is that CBR is an incremental approach. This means that each time a problem is solved, this new experience is retained, making it immediately available for future problems [12].

The processes involved in CBR can be represented by a schematic cycle (Figure 1), which is comprised of the tasks of retrieving the most similar case, reuse/adapt the case to try to resolve the problem, revise the proposed solution if necessary, and retain new solution as part of a new case.

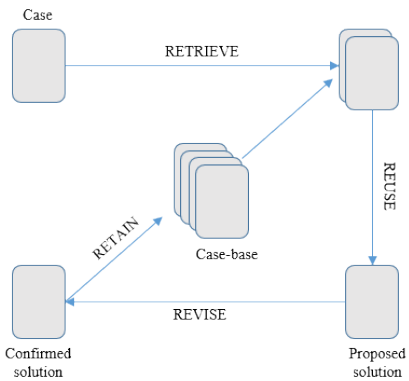


Figure 1 – CBR cycle, introduced by Aamodt and Plaza [11]

CBR can be integrated with other techniques. Marling et al. [13] present CBR integration with RBR and constraint-satisfaction problem (CSP) solving. Furthermore, they discuss CBR integration with model-based reasoning (MBR), genetic algorithms, and information retrieval. RBR was the first modality to be successfully integrated with CBR [13].

Rule-based reasoning

RBR is a methodology whose representation of knowledge is in the form of IF–THEN rule statements. Rules are patterns, so the RBR engine searches for patterns in the rules that match patterns in the data. RBR is an ideal approach for solving simple problems in which there are few rules [14]. In RBR, the problem solving complexity is directly proportional to the number of rules necessary to match the pattern of data. Furthermore, RBR lacks the ability to learn due to the difficulty of acquiring new expertise in pattern matching or new rules [14].

The basic form of a rule is the following:

```
IF <conditions>
THEN <conclusion>
```

where <conditions> represents the rule conditions, and can be connected by logical operators such as AND, OR, NOT, etc., forming a logical function. When rule conditions are satisfied, the <conclusion> is derived and the rule is said to "trigger" [9].

Methods

In medical decision support systems, the use of CBR or RBR methodologies is common [15]. In the proposed approach, we used CBR as the main reasoning process, and RBR was used to improve part of this process. The idea is that our approach can be used in a system that assists the physician in the early diagnosis of cancer. During a medical appointment, the patient tells the doctor some personal data and the symptoms that he/she is feeling. The physician will add this information to the system along with the signals perceived by the patient. The system will search in the database for the most similar case to that of the patient. Based on this result, the doctor may state the prognosis, and request tests to confirm the presence or absence of disease (Figure 2). The proposed model may act both as a decision support system for less experienced clinicians and as a second opinion for experts.

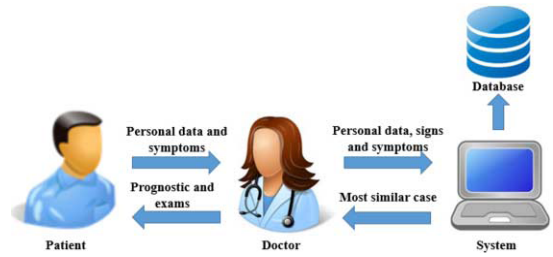


Figure 2 – System representation

Our methodology is composed of four main steps: data collection, case representation, similarity measures definition and rules definition (Figure 3). The first step is necessary for both the CBR application and also the RBR. The second and third steps correspond to two basic elements of a CBR system [12] and the fourth stage corresponds to RBR methodology. We applied RBR to define the case's attribute weights, used in the global similarity function.

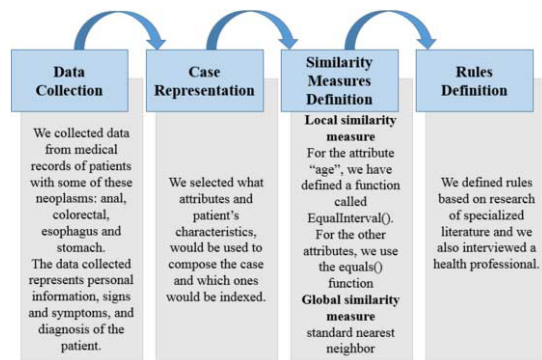


Figure 3 – Main steps of the methodology

In the following subsections, we present in more detail the four steps of our methodology.

Data Collection

With the approval of the Napoleão Laureano Hospital ethics committee, we collected data from the medical records of patients. We kept the privacy of their data. We collected the medical records ID from the agenda of the specialist in gastrointestinal cancer and we focused, more specifically, on patients with the following gastric neoplasms: anal, colorectal, esophagus and stomach.

The data collected corresponds to some personal information such as age, family history, signs and symptoms such as cutaneo-mucosal pallor and dysphagia, respectively, and the diagnosis of the patient. It was not possible to collect all the data we would have liked because some records had little information about the signs and symptoms of the patient. Furthermore, many of them had little digitalized infomation, hindering our collection process.

In this research, we used forty-eight cases from real patients: six cases of patients with anal cancer, six with esophageal cancer, fifteen with colorectal cancer, and twenty-one with stomach cancer.

Case Representation

To represent the cases, we used a set of [attribute – value]. An [attribute – value] system is a basic knowledge representation framework comprising a table with columns representing attributes and rows representing objects. Each table cell therefore designates the value (also known as the "state") of a particular attribute of a particular object. In this work, each object is a case and can be represented by a problem, personal data, signs and symptoms, and also by a solution, the diagnosis (Figure 4).

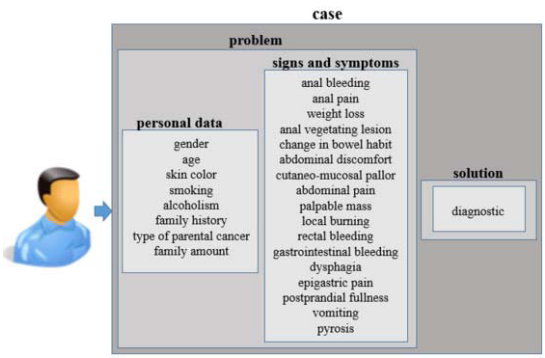


Figure 4 – Representation of the case in the proposed approach

In our approach, a case consists of twenty-six attributes, most of them boolean (i.e., they assume true or false values). Types and the respective values of the remaining attributes can be viewed in Table 1.

Table 1 – Attributes Configuration

Attribute	Type	Value
gender	String	male / female
age	Integer	1 to 110
skin color	String	white / brown / black
smoking	String	non-smoking / smoking / ex-smoking
alcoholism[4]	String	non-alcoholic / alcoholic / ex-alcoholic
family history	String	none / 1º degree / 2º degree / 3º degree
type of parental cancer	String	none / mouth / colorectal / stomach / eye / ovary / lung
family amount	String	0 / 1 / >1

Similarity Function Definition

Each attribute in a case is a piece of information about the case. An important attribute for the recovery process is called the "attribute index". Each index has a set weight, which represents the importance of that attribute in the recovery process, and is typically instantiated by the user with a value between 1 and 10.

The recovery process is based on a similarity function. In this work, we used the standard nearest neighbor method [12]:

$$sim(Q, C) = \frac{\sum_{i=1}^{n} f(Qi,Ci) x\, wi}{\sum_{i=1}^{n} wi} \quad (1)$$

This function returns the global similarity value between two cases, Q and C. Qi corresponds to the attribute value i of a new case and Ci corresponds to the same attribute i of the case recovered from the case-base. Wi is the weight of i attribute and n is the number of attributes of a case. Thus, n is equal to twenty-three, since the attributes "family history," "number of family", and "prognosis" are not used in the calculation of the equation.

In (1), "f" is the function of local similarity of each attribute. All of them use the local similarity function Equal(), except for the age attribute. For this, a new function called EqualInterval(), which calculates the similarity of the attribute according to age ranges (<= 60, 61-70 and >= 71) defined by us. Thus, for this attribute, the values of 74 and 80 are coded identically.

The twenty-three features used in the calculation of similarity have a default weight of "1". Some attributes have altered weights according to pre-established rules mentioned in the later section. We believe that the weights in general should be chosen by a specialist in gastrointestinal cancer in order to increase the accuracy of the system.

As we have a small case-base, the recovery method that we use is the sequential, that is, the measure of similarity is calculated for all cases of the base [12].

Rules Definition

The rules created and used in our approach were based on information extracted from a medical book [10] and the National Cancer Institute[5] (INCA) website. The weights used in the rules were decided with the help of a general practitioner. We interviewed the general practitioner and used a template to formulate the questions: "How important is the X attribute to diagnose cancer Y?", where X corresponds to the some case attribute and Y to some type of cancer. The answers were collected on a scale from 1 to 10.

1st rule: This corresponds to the patient's family history. Patients who have a relative who has or had a particular type of cancer is more likely to also have cancer. Among the types of cancer studied in this research, only colorectal and stomach cancer consider family history.

Variables	Condition	Action
type_ca_parental	type_ca_parental = colorrectal OR type_ca_parental = stomach	type_ca_parental.weight ← 5

Figure 5 – First Rule

2nd rule: This also corresponds to the patient's family history. If the number of a patient's relatives who have cancer is greater than 1, then the patient will be more likely to also have this neoplasm.

Variables	Condition	Action
amount_family type_ca_parental	amount_family > 1	type_ca_parental.weight ← 8

Figure 6 – Second Rule

3rd rule: Dysphagia is the main clinical manifestation that occurrs in patients with esophageal cancer.

[4] a chronic disorder characterized by dependence on alcohol

[5] National Cancer Institute http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home

Variables	Condition	Action
dysphagia	dysphagia = true	dysphagia.weight ← 5

Figure 7 – Third Rule

4th rule: According to the data from the literature [10], weight loss is a common feature in patients with colorectal, esophagus, or stomach cancers. However, there are cases of patients with anal cancer that also presented with weight loss.

Variables	Condition	Action
weight_loss	weight_loss = true	weight_loss.weight ← 2

Figure 8 – Fourth Rule

5th rule: Postprandial fullness is a feeling of stomach fullness, and is characteristic of patients with stomach cancer.

Variables	Condition	Action
post_prandial_fullness	post_prandial_fullness = true	post_prandial_fullness.weight ← 5

Figure 9 – Fifth Rule

6th rule: Anal bleeding is the most common feature present in patients with anal cancer [10]. However, patients with colorectal cancer may also exhibit this symptom.

Variables	Condition	Action
anal_bleeding	anal_bleeding = true	anal_bleeding.weight ← 5

Figure 10 – Sixth Rule

7th rule: Anal pain is a feature of patients with anal cancer, but it also appears as a symptom in patients with colorectal cancer.

Variables	Condition	Action
anal_pain	anal_pain = true	anal_pain.weight ← 5

Figure 11 – Seventh Rule

8th rule: A change in bowel habits (diarrhea or constipation) is a warning sign for colorectal cancer. There are cases of patients with stomach cancer who also have this symptom.

Variables	Condition	Action
change_in_bowel_habit	change_in_bowel_habit = true	change_in_bowel_habit.weight ← 5

Figure 12– Eighth Rule

9th rule: Abdominal pain and changes in bowel habits are symptoms of colorectal cancer. They are also often present in patients with stomach cancer.

Variables	Condition	Action
abdominal_pain	abdominal_pain = true	abdominal_pain.weight ← 3

Figure 13 – Ninth Rule

Validation Method and Results

To validate the system, we developed a prototype, and we used k-fold cross-validation [16]. As we have forty-eight cases and the number of cases for each neoplasm is a multiple of three, we split the data into three blocks, each one with sixteen cases.

We used each block as input to the system. First, we made tests without the use of rules (i.e., all attributes independent of the value had a weight of "1"). Then, we redid the tests using the rules previously mentioned. The result of the three iterations can be seen in Table 2, Table 3, and Table 4, respectively.

The diagnosis accuracies by fold and by type of cancer are shown in Table 5 and Table 6, respectively.

Table 2 – Results of the first test block

Nº	Real diagnosis	Without rules		With rules	
0	----------	1° case	hit	1° case	hit
1	**Anal**	anal	1	anal	1
2	**Anal**	anal	1	anal	1
3	**Colorectal**	esophagus	0	colorectal	1
4	**Colorectal**	colorectal	1	colorectal	1
5	**Colorectal**	esophagus	0	esophagus	0
6	**Colorectal**	anal	0	anal	0
7	**Colorectal**	stomach	0	colorectal	1
8	**Esophagus**	esophagus	1	esophagus	1
9	**Esophagus**	esophagus	1	esophagus	1
10	**Stomach**	stomach	1	stomach	1
11	**Stomach**	stomach	1	stomach	1
12	**Stomach**	colorectal	0	colorectal	0
13	**Stomach**	stomach	1	stomach	1
14	**Stomach**	stomach	1	stomach	1
15	**Stomach**	stomach	1	stomach	1
16	**Stomach**	stomach	1	stomach	1
sum	----------	----------	11	----------	13

Table 3 – Results of the second test block

Nº	Real diagnosis	Without rules		With rules	
0	----------	1° case	hit	1° case	hit
1	**Anal**	anal	1	anal	1
2	**Anal**	anal	1	anal	1
3	**Colorectal**	anal	0	anal	0
4	**Colorectal**	stomach	0	colorectal	1
5	**Colorectal**	colorectal	1	colorectal	1
6	**Colorectal**	anal	0	anal	0
7	**Colorectal**	stomach	0	stomach	0
8	**Esophagus**	stomach	0	stomach	0
9	**Esophagus**	esophagus	1	esophagus	1
10	**Stomach**	esophagus	0	stomach	1
11	**Stomach**	colorectal	0	stomach	1
12	**Stomach**	stomach	1	colorectal	0
13	**Stomach**	stomach	1	stomach	1
14	**Stomach**	colorectal	0	stomach	1
15	**Stomach**	stomach	1	stomach	1
16	**Stomach**	esophagus	0	stomach	1
sum	----------	----------	7	----------	11

Discussion and Conclusion

Many researchers have developed different approaches to predict, diagnose, and classify cancers, but in general, only a single type of cancer is discussed. In this research, we focused on four types of gastrointestinal cancer. In addition, we used rules to customize the cases. The diagnosis accuracies by fold

and by type of cancer are shown in Table 5 and Table 6, respectively.

To assess if our approach increased the diagnosis accuracy compared to using only the CBR approach, we used the paired t-test with 95% confidence interval and got p-value = 0.02664, refuting the null hypothesis (H0 = The CBR performance is the same as the proposed hybrid approach). Given this, we confirmed our expectations.

The limitations of this study are related to the quantity and quality of the cases and weights. In addition, we believe that with the help of an oncologist, we could improve the rules and attribute weights. Even though the model training technique used data from a specific population group, the cross-validation results might not be enough to generate adequate data for a reliable model.

In future works, we will extend the case-base, and will seek help from a medical expert to validate the rules and the weights associated with them. Furthermore, we will discuss the remaining phases of the CBR cycle.

Table 4 – Results of the third test block

Nº	**Real diagnosis**	**Without rules**		**With rules**	
0	----------	**1° case**	**hit**	**1° case**	**hit**
1	**Anal**	colorectal	0	anal	1
2	**Anal**	colorectal	0	anal	1
3	**Colorectal**	stomach	0	stomach	0
4	**Colorectal**	colorectal	1	colorectal	1
5	**Colorectal**	stomach	0	anal	0
6	**Colorectal**	colorectal	1	colorectal	1
7	**Colorectal**	stomach	0	stomach	0
8	**Esophagus**	colorectal	0	esophagus	1
9	**Esophagus**	stomach	0	esophagus	1
10	**Stomach**	esophagus	0	esophagus	0
11	**Stomach**	colorectal	0	stomach	1
12	**Stomach**	stomach	1	stomach	1
13	**Stomach**	stomach	1	stomach	1
14	**Stomach**	stomach	1	stomach	1
15	**Stomach**	stomach	1	stomach	1
16	**Stomach**	stomach	1	stomach	1
sum	----------	----------	7	----------	12

Table 5 – Diagnosis accuracies by fold

	Folds			
	First	**Second**	**Third**	**Mean**
Without rules	68.75%	43.75%	43.75%	52.08%
With rules	81.25%	68.75%	75%	75%
Gain	12.50%	25%	31.25%	22.92%

Table 6 – Diagnosis accuracies by type of cancer

	Type of cancer			
	Anal	**Colorectal**	**Esophagus**	**Stomach**
Without rules	66.66%	26.66%	50%	66.66%
With rules	100%	46.66%	83.33%	85.71%
Gain	33.34%	20%	33.33%	19.05%

Acknowledgments

We would like to thank CAPES and Federal Institute of Paraiba for the financial support. We would like to extend our thanks to employees and friends of Napoleao Laureano Hospital.

References

[1] Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine 2001 June: 7: 673-679.

[2] Lin RH. An intelligent model for liver disease diagnosis. Artificial Intelligence in Medicine 2009 May: 47: p. 53—62.

[3] Hsu KH, Chiu C, Chiu NH, Lee PC, Chiu WK, Liu TH, et al. A case-based classifier for hypertension detection. Knowledge-Based Systems 2011 July: 24: 33-39.

[4] Chuang CL. Case-based reasoning support for liver disease diagnosis. Artificial Intelligence in Medicine 2011 June: 53: 15-23.

[5] Salem ABM, and El Bagoury BM. A Case-Based Adaptation Model for Thyroid Cancer Diagnosis Using Neural Networks. In Society, Florida Artificial Intelligence Research. Cairo: Egypt, 2003; pp. 155-9.

[6] Zubi ZS, and Saad RA. Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer. In Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Base. Tripoli: Libya, 2011; pp. 32-7.

[7] Keles A, Keles A, and Yavuz U. Expert system based on neuro-fuzzy rules for diagnosis breast cancer. Expert Systems with Applications 2011: 38: 5719–26.

[8] Sharaf-elDeen DA, and Moawad IF, E. MK. A Breast Cancer Diagnosis System using Hybrid Case-based Approach. International Journal of Computer Applications 2013 June: 72: 14-9.

[9] Prentzas J, and Hatzilygeroudis I. Categorizing approaches combining rule-based and case-based reasoning. Expert Systems 2007 May: 24: 97-122.

[10] Engel CL. Tumores Gastrointestinais Writers EM, ed., 2010.

[11] Aamodt A, and Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications 1994 March: 7: 39-59.

[12] Wangenheim GV, and Wangenheim V. Raciocínio Baseado em Casos Ltda EM, ed., 2003.

[13] Marling C, Sqalli M, Rissland E, Muñoz-Avila H, and Aha D. Case-Based Reasoning Integrations. AI Magazine 2002: 23: 69-86.

[14] Lim TP, Husain W, and Zakaria N. Recommender System for Personalised Wellness Therapy. International Journal of Advanced Computer Science and Applications 2013: 4: 54-60.

[15] Rossille D, Laurent JF, and Burgun A. Modelling a decision-support system for oncology using rule-based and case-based reasoning methodologies. International Journal of Medical Informatics 2005 June: 74: p. 299—306.

[16] Refaeilzadeh P, Tang L, and Liu H. Cross-Validation: Springer US, 2009.

Address for correspondence:

Renata M. Saraiva renatams@copin.ufcg.edu.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-867

Identifying Clinical Study Types from PubMed Metadata: The Active (Machine) Learning Approach

Adam G. Dunn [a], Diana Arachi [a], Florence T. Bourgeois [b,c]

[a] *Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW, Australia*
[b] *Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA, USA*
[c] *Department of Pediatrics, Harvard Medical School, Boston, MA, USA*

Abstract

We examined a process for automating the classification of articles in MEDLINE aimed at minimising manual effort without sacrificing accuracy. From 22,808 articles pertaining to 19 antidepressants, 1000 were randomly selected and manually labelled according to article type (including, randomised controlled trials, editorials, etc.). We applied a machine learning approach termed 'active learning', where the learner (machine) selects the order in which the user (human) labels examples. Via simulation, we determined the number of articles a user needed to label to produce a classifier with at least 95% recall and 90% precision in three scenarios related to evidence synthesis. We found that the active learning process reduced the number of training instances required by 70%, 19%, and 14% in the three scenarios. The results show that the active learning method may be used in some scenarios to produce accurate classifiers that meet the needs of evidence synthesis tasks and reduce manual effort.

Keywords:

Machine Learning; Databases, Bibliographic; Antidepressants.

Introduction

When undertaking reviews or synthesis of clinical evidence, researchers are faced with the task of identifying all relevant published studies in bibliographic databases. This has become increasingly difficult and time-consuming with the dramatic increase in the number of studies being published each year.

To improve the efficiency of evidence synthesis tasks that require the comprehensive identification of specific types of articles, researchers have analytically devised search queries or filters, and applied them to large bibliographic databases such as MEDLINE [1-8]. Others have proposed methods for context-specific searches by partially or completely replacing the need for manual screening and classification using machine learning methods [9-11].

In an effort to maximise the proportion of relevant articles identified (recall) in these searches – which ideally should be 100% for systematic reviews – researchers are often forced to sacrifice the proportion of irrelevant articles also included as relevant (precision), leaving them with many articles that require further manual query and classification.

Here we consider the application of machine learning to the task of article classification, and propose the use of *active learning* [12], an approach aimed at reducing the number of examples that a user (human) needs to provide to the learner (machine). Active learning has been used to classify clinical studies for systematic reviews previously [11], and for well over a decade in biomedical research applications ranging from drug discovery to biomedical informatics [13, 14].

Our aim was to examine whether active learning could be used to reduce the manual workload for classifying articles in tasks without substantially reducing the quality of the process. To do this, we conducted experiments for three scenarios of evidence synthesis, using MEDLINE metadata on a large corpus of articles about antidepressants.

Methods

Data

We used PubMed to search MEDLINE on 9 December 2014 for articles that included in the title, abstract, or MeSH terms, the names or synonyms of one or more of the 19 antidepressants that were approved for use by the US Food and Drug Association (FDA) between 1995 and 2005. The search returned 40,098 unique articles. From this set, we selected the 22,808 articles for which the antidepressant of interest was already approved by the FDA when the article was published, and then randomly selected 1000 for inclusion in the study. We selected antidepressants for a study of author influence in postapproval opinion/commentary.

Two investigators (AD and DA) classified each article as one of eight article types: (1) systematic review; (2) non-systematic review and meta-analysis; (3) guideline; (4) randomised controlled trial; (5) other clinical study (e.g. non-randomised trial, cost-benefit analysis); (6) case study; (7) non-clinical study (animal models, chemistry for synthesis, testing, or where antidepressants are not the primary topic); and (8) comment or opinion piece. Agreement between the two investigators was 90% (Cohen's kappa = 0.87) and disagreements were resolved by discussion.

Specific characteristics (metadata) were abstracted from the articles to form the set of potential features, which are used as the basis for training the machine learning classifiers. These metadata included article titles and abstracts, MeSH terms, and publication types assigned in MEDLINE. For titles and abstracts, each word with more than three letters was extracted and binary features created to represent their presence or absence in each article. MeSH terms and publication types were used to create binary features without change. No other pre-processing methods (such as stemming or the use of stop words) were applied to the metadata prior to their use in the classifier training. The total number of features that could be used to train the classifiers was 5,606.

Three scenarios were created as examples of evidence review, synthesis, and surveillance tasks that might benefit from automation. In Scenario 1, the aim was to find all clinical trials for use in a systematic review (4 and 5, above). In

Scenario 2, the aim was to identify all randomised controlled trials for inclusion in a meta-analysis (4, above). In Scenario 3, the aim was to find only opinion pieces, commentaries, and non-systematic reviews (2 and 8, above) for examining the distribution of contributions by individual authors.

The active learning process

We defined the ordered set of articles that the user is asked to label by the learner as the *labelling queue*. The unique feature of the active learning process is that the learner iteratively trains new classifiers each time it receives an additional label from the user, and uses the resulting information to determine the order of the labelling queue.

The aim of the learner is to minimise the length of the labelling queue by converging more quickly on a classifier that is able to exceed the performance requirements.

The proposed process for applying active learning to this task includes two phases (an alternative to a previous approach [11]): the first is for when there are too few examples in each class to examine the distribution of features across the two classes and identify meaningful features (Phase I); and the second when a new classifier is trained in each step and used to select the next article to place in the labelling queue (Phase II).

Phase I

In Phase I, the issue is that there is not enough information available for the learner to know enough about the distribution of features across relevant/irrelevant articles, and thus not enough identifiable features capable of discriminating between the two classes to effectively construct a classifier.

A simple solution in this stage is to progress at random until enough useful information is available for the active learning approach. To do this, we take advantage of the statistical distribution of features among known articles. The user is asked to label articles selected at random until there are a minimum of five features that exhibit a significant difference ($p<0.05$ in a Fisher's exact test) between the classes. These features then become available for the classifiers constructed in the second phase.

Alternative methods for finding useful articles at this stage involve clustering the articles based on the distribution of features, without knowing to which class any of the articles belong [15]. We did not consider these alternatives in this study but these could reduce the workload in Phase I.

Phase II

Phase II is the active learning phase. In each round, a new classifier is trained using the labels from the labelling queue, and the information produced by applying the new classifier to the pool of remaining unlabelled articles is used to determine the next article to be labelled by the user.

For selecting useful features, we applied a simple statistical method in which features from all labelled articles (so far) are compared across the two classes (either relevant or irrelevant articles). In the statistical testing, Fisher's exact tests were applied to each feature and those with the lowest p-values were then included as features to train the classifiers. In this study, we chose to select the features that remained after a Bonferroni correction was applied, or the five features with the lowest p-values, whichever was larger.

We used support vector machines as the machine learning algorithm in these experiments, as it is appropriate for tasks such as document classification [16]. The support vector machines were constructed using linear kernels and a least squares method. To apply the active learning approach to support vector machines, the learner selects the unlabelled article that is closest to the hyperplane found to best separate the relevant and irrelevant articles during training, and passes that article to the user for labelling [12, 17].

Note that we only considered one approach for feature selection, one type of machine learning algorithm, and applied the rules consistently across all simulations. Alternative choices for feature selection and for classifiers could have affected the performance of the active learning and passive learning approaches in different ways.

Analysis

We evaluated the active learning process by estimating the risk of not achieving a pre-specified performance for each number of labelled articles. The pre-specified performance criteria were 95% recall (the percentage of relevant articles that were identified), and 90% precision (the percentage of identified articles that were relevant). Using these measures, we compared the results of the active learning simulations to a passive learning baseline, in which the articles labelled by the user are selected at random from the pool of all articles. This is equivalent to a process that continues with Phase I and never progresses to Phase II.

To estimate the risk of imperfect recall at a given labelling queue length, we ran repeated simulations to determine how many articles would need to be labelled by the user before a classifier with 95% recall and 90% precision was trained. In each case, 500 articles were selected (by stratified random sampling) for training and the remaining 500 were used as a holdout test set (which differs from a previous approach [11]). The performance results were determined by applying the classifer to the holdout set. For each of the three scenarios, we ran 1000 simulations for the passive learning baseline, re-sampling 500 articles each time.

The features that were most often used to train the classifiers that first reached 95% recall and 90% precision were reported in the active learning process as 'positive' features if they were more often present in the relevant articles, and 'negative' features otherwise.

We also examined how the process might work in practice, by conducting simulations in which the user labels 100 articles by active learning or the passive learning baseline. In these experiments, we simulated the approach 1000 times using 500 articles, but we determined the performance by testing it using the 500 articles (labelled and unlabelled) to find out if labelling 20% of articles would ensure a reliable classifier.

Results

Non-clinical studies made up the greatest proportion in the sample of 1000 articles, followed by other clinical studies, randomised controlled trials, case studies and non-systematic reviews (Table 1).

Table 1– Distribution of study types

Article Type	Frequency	Proportion
Non-clinical study	410	41%
Other clinical study	201	20%
Randomised controlled trial	130	13%
Case study	105	10%
Non-systematic review	101	10%
Systematic review	30	3.0%
Opinion piece/commentary	22	2.2%
Guideline	1	0.1%
Total	1000	100%

Scenario 1: clinical studies for use in systematic reviews

In Scenario 1, the task was to identify the 331 randomised controlled trials and other clinical studies (4 and 5, above). Applying the active learning process, the median number of articles labelled by the user when the classifier first exceeded 95% recall and 90% precision was 108 (IQR 78-146). To reach the same level of performance, the passive learning process required a median of 366 articles (IQR 284-460). This corresponds to a workload reduction of 70% in the number of labelled articles required to train a classifier (Figure 1). The results also show that for a small proportion of active learning simulations, the classifier never reached 95% recall and 90% precision (even after the user had labelled 200 articles), while for the passive learning simulations, the risk of training a poor classifier for unseen articles remained even after the user had labelled the entire training set of 500 articles.

The positive features most commonly included in the final classifiers generated by the active learning process included the MeSH terms "female" (97% of final classifiers) and "adult" (96%), the publication types "Randomized Controlled Trial" (99%) and "Clinical Trial" (97%), and abstract words "study" (70%) and "patients" (47%). Negative features included MeSH terms "Animals" (100%) or "rats" (49%), and publication type "Review" (94%) or "Case Reports" (86%).

To examine how the active learning process might work in practice, we measured the performance of the classifiers after the user had labelled 100 articles. The performance is considered *in sample*, measuring the recall, precision, and F_1-score of the 500 sampled articles including those that were already labelled (as existing methods have done [11]). The active learning process produced a median recall of 93%, median precision 96%, and median F_1-score 0.94. In the passive learning process, the median recall was 90%, median precision 90%, and median F_1-score 0.90 (Table 2). The differences in recall, precision, and F_1-score were significant ($p<0.001$) under a two sample Kolmogorov-Smirnov test.

Scenario 2: randomised controlled trials for meta-analyses

In Scenario 2, the median number of articles labelled in the passive learning process was 28 (IQR 22-38) compared to 24 (IQR 20-30) in the active learning process, representing a 14% reduction in workload (Figure 2). The results likely reflect the standardised ways in which randomised controlled trials are published, rather than the unbalanced nature of the dataset (130 relevant articles). Among the positive features that were most often used by the classifiers in the active learning process was the publication types "Randomized Controlled Trial" (99%), abstract terms "blind" (64%), "placebo" (62%), "double" (60%), and "week" (21%), and the MeSH terms "DoubleBlind_Method" (61%) and "Adult" (38%).

The marginal difference between the two approaches is reflected in the results of the practical test after the user has labelled 100 articles (Table 2), where both active and passive methods produce similar results. The differences in recall, precision, and F_1-score were significant ($p<0.001$).

Scenario 3: opinion pieces and non-systematic reviews for a task in pharmacosurveillance

In Scenario 3, the median number of articles labelled in the passive learning process was 42 (IQR 30-58) and in the active learning process 52 (IQR 37-79). This represents a 19% workload reduction in the number of training instances required (Figure 3). The results in Scenario 3 also show that a small proportion of the passive learning classifiers never reached the pre-specified stopping criteria, while the risk was minimal in the active learning process after 150 articles had been labelled. The positive features that were most often used by the classifiers in this scenario were the abstract words "therapy" (25%), "trials" (21%), "agents" (23%), "effective" (17%), and the publication type "Review" (95%). The negative features included abstract words "were" (48%) and "study" (27%), and the MeSH term "Male" (34%).

Reflecting the small gap between the passive and active learning processes, the differences in performance after the user labelled 100 articles were also relatively small (Table 2). The differences in recall, precision, and F_1-score were significant ($p<0.001$).

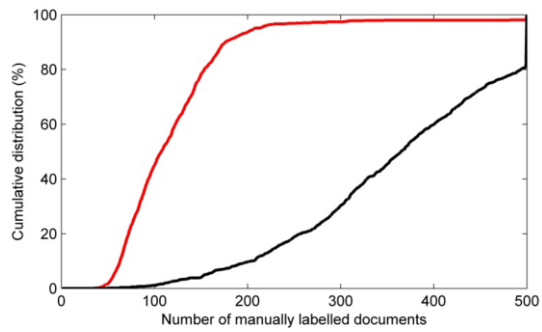


Figure 1– Clincial studies (S1): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).

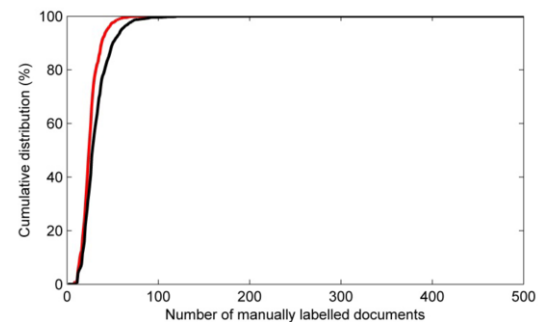


Figure 2– Randomised Controlled Trials (S2): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).

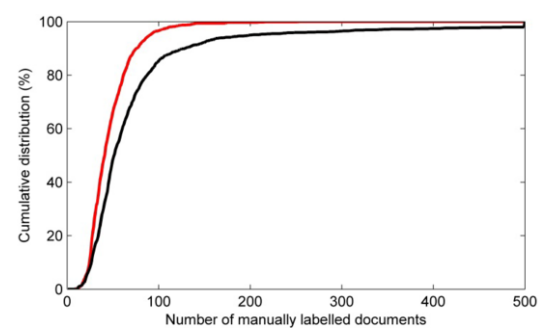


Figure 3– Opinions & Commentaries (S3): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).

Table 2– Comparing the performance (in the training set) after 100 user labellings to examine the process in practice.

Process and scenario	Median recall (interquartile range)	Median precision (interquartile range)	Median F_1-score (interquartile range)
Passive learning process			
S1: All clinical trials	0.90 (0.88-0.93)	0.90 (0.88-0.91)	0.90 (0.89-0.91)
S2: Randomised controlled trials	0.98 (0.97-0.98)	0.99 (0.98-0.99)	0.98 (0.97-0.99)
S3: Non-systematic reviews and opinion pieces	0.95 (0.93-0.96)	0.94 (0.92-0.96)	0.94 (0.93-0.95)
Active learning process			
S1: All clinical trials	0.93 (0.89-0.95)	0.96 (0.94-0.97)	0.94 (0.92-0.95)
S2: Randomised controlled trials	0.98 (0.98-0.99)	0.99 (0.99-0.99)	0.99 (0.99-0.99)
S3: Non-systematic reviews and opinion pieces	0.97 (0.96-0.97)	0.97 (0.96-0.98)	0.97 (0.96-0.97)

Discussion

We applied an active learning method to the article classification task, demonstrating that the approach can reduce the need for training instances without sacrificing recall in some scenarios. The results showed a reduction of 70%, 14%, and 19% in the number of training instances required to produce 95% recall and 90% accuracy in three evidence synthesis scenarios. These results suggest that active learning may be a feasible solution for article classification in some circumstances.

In previous methods for classifying articles in MEDLINE that did not employ machine learning methods, researchers analytically derived general seearch terms by examining very large numbers of initial articles [1-8]. These methods were generally unable to identify every relevant article, and recall could only be increased by sacrificing precision, often resulting in article lists in which fewer than half of the articles returned were relevant. As a result, extensive manual review would be required in a second stage of screening. It is likely that the active learning process presented here would improve on these approaches because it uses fewer labels to provide a low risk of missing a relevant study without sacrificing precision.

Machine learning approaches used to classify articles from MEDLINE and other bibliographic databases have been proposed for different purposes [10, 18], and active learning has been proposed for identifying relevant articles to include in systematic reviews [11]. In one example, authors used 10,000 articles as a training set, achieving 73.7% precision and 61.5% recall when identifying articles that were scientifically rigorous [9]. In another, the aim was not to label studies, but rather to order the returned documents so that the most relevant articles were listed first [19]. In the study that applied active learning, the performance of the method was determined by describing a measure of workload reduction, and the results indicated the potential for a 40-50% reduction without sacrificing perfect recall [11].

Another group of studies considers a more general task: the automatic assignment of MeSH labels to articles [20-25]. This is similar to the tasks we considered here because an accurate mapping of MeSH labels to articles could be used to replace or complement a context-specific screening of article types for inclusion in an evidence synthesis task.

The main contribution present in this study is the application of the active learning approach to article classification for evidence synthesis tasks, and an evaluation of the approach using a holdout set, examining the risk of not meeting pre-defined performance criteria for a given number of labellings. Compared to other applications of machine learning, our approach considers a relatively simple problem with binary classification by document type. However, it would be simple to extend the approach to consider more complex screening requirements where only specific study designs were relevant (e.g. specific comparators or outcome measures). Our study is also relatively simple because it considers one form of feature construction (no stemming, stop words, or n-grams), only one type of machine learning algorithm (support vector machines, no stacking or boosting), and a simple method for selecting features (selecting from a list ranked by p-values). The approach may be improved by introducing more sophisticated machine learning methods.

The differences between the three scenarios suggest that class imbalance (for example, where very few articles are relevant), and the context of the task (which influences the features that may be suitable) may both influence the length of the labelling queue required to produce high enough levels of recall and precision to make the approach worthwhile. This means that tasks aimed at identifying rare article groups may benefit least from the active learning approach but this is speculation because we considered only one scenario in which the classes were heavily imbalanced. However, identifying rare types of articles using the active learning approach has been considered in depth elsewhere [17], warranting further investigation.

Impossible guarantees and empirical estimates of risk

Tasks in evidence synthesis often require 100% recall in screening tasks [26], so it is worth considering whether an automated method can be developed to guarantee perfect recall. In inductive learning, the user and the learner are unable to calculate how well the trained classifiers will perform on unseen examples. This means that although we were able to stop the simulations when the classifier met the pre-specified performance criteria in the experiments here, this is not possible in practice.

This problem is exacerbated by the variability in the classifiers being produced using both passive and active learning within and across the scenarios. This means that the number of articles that need to be labelled by the user to meet a specific risk level may vary for different tasks. While the risk of missing relevant articles is reduced by the active learning process, the variability suggests that we do not yet have a way to (analytically or empirically) determine how many articles need to be labelled to guarantee an acceptable level of risk.

Limitations

We used only one type of classifier and one method for selecting and constructing features. While support vector machines are appropriate for tasks with a large number of features and document classification tasks in particular, other classifiers and methods for selecting features could affect the results in unexpected ways. Equally, the stopping criteria for Phase I and the final performance criteria that we reported on could have been chosen differently, and these choices could also affect the results.

We examined only one class of clinical interventions (antidepressants), and three relatively simple scenarios in

evidence synthesis. While it would be reasonable to expect that other drug classes and conditions would have similar distributions for article types and produce the same levels of performance, this may not hold for atypical interventions or conditions. Specific search queries aimed at identifying studies that meet specific design specifications may not produce the same results. For example, it may be more difficult to construct a classifier that can separate from all articles those that report on clinical trials comparing two specific interventions and report a specific measurable outcome. Other than issues with class imbalance for cases where there are very few relevant articles, the approach is general so we have no reasons to expect that the method would perform any differently in these cases.

Conclusion

In this study, we considered the problem of article classification as a shared task between a (human) user and a (machine) learner. Active learning may provide the potential to reduce the manual effort required to classify articles for context-specific tasks in evidence synthesis without reducing the performance. However, variability in the performance of classifiers within and across the scenarios suggests that we do not yet have a way to both ensure that no relevant articles are missed and substantially reduce the workload required to screen documents for inclusion in a study.

References

[1] Haynes R, McKibbon K, Wilczynski N, Walter S, Werre S, Hedges Team, Optimal search strategies for retrieving scientifically strong studies of treatment from medline: Analytical survey. BMJ, 2005 **330**:1179.

[2] Haynes R, Wilczynski N, McKibbon K, Walker C, Sinclair J, Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc, 1994 **1**:447-58.

[3] Haynes R, Wilczynski N, Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. BMJ, 2004 **328**:1040.

[4] Montori V, Wilczynski N, Morgan D, Haynes R, Optimal search strategies for retrieving systematic reviews from medline: Analytical survey. BMJ, 2005 **330**:68.

[5] Robinson K, Dickersin K, Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol, 2002 **31**:150-3.

[6] Shojania K, Bero L, Taking advantage of the explosion of systematic reviews: An efficient medline search strategy. Eff Clin Pract, 2000 **4**:157-62.

[7] Wilczynski N, Haynes R, Hedges Team, Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. BMC Med, 2004. **2**:23.

[8] Wilczynski N, Haynes R, Lavis J, Ramkissoonsingh R, Arnold-Oatley A, Hedges Team, Optimal search strategies for detecting health services research studies in MEDLINE. CMAJ, 2004. **171**:1179-85.

[9] Kilicoglu H, Demner-Fushman D, Rindflesch T, Wilczynski N, Haynes B, Towards automatic recognition of scientifically rigorous clinical research evidence. J Am Med Inform Assoc, 2009 **16**:25-31.

[10] Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis C, Text categorization models for high-quality article retrieval in internal medicine. J Am Med Inform Assoc, 2005 **12**:207-16.

[11] Wallace B, Trikalinos T, Lau J, Brodley C, Schmid C, Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics, 2010 **11**:55.

[12] Tong S, Koller D, Support vector machine active learning with applications to text classification. J Mach Learn Res, 2002 **2**:45-66.

[13] Warmuth M, Liao J, Rätsch G, Mathieson M, Putta S, Lemmen C, Active learning with support vector machines in the drug discovery process. J Chem Inf Comput Sci, 2003 **43**:667-73.

[14] Liu Y, Active learning with support vector machine applied to gene expression data for cancer classification. J Chem Inf Comput Sci, 2004 **44**:1936-41.

[15] Nguyen H, Smeulders A, Active learning using pre-clustering. In: Proceedings of the 21st International Conference on Machine Learning, ACM. Banff, Canada 2004 623-30.

[16] Joachims T, Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of the 10th European Conference on Machine Learning. Chemnitz, Germany 1998 137-42.

[17] Ertekin S, Huang J, Bottou L, Giles L, Learning on the border: Active learning in imbalanced data classification. In: Proceedings of the 16th Conference on Conference on Information and Knowledge Management, ACM. Lisbon, Portugal 2007 127-36.

[18] Sebastiani F, Machine learning in automated text categorization. ACM Comput Surv, 2002 **34**:1-47.

[19] Cohen A, Optimizing feature representation for automated systematic review work prioritization. AMIA Annu Symp Proc, 2008 **2008**:121-5.

[20] Trieschnigg D, Pezik P, Lee V, de Jong F, Kraaij W, Rebholz-Schuhmann D, MeSH up: effective MeSH text classification for improved document retrieval. Bioinformatics, 2009 **25**:1412-18.

[21] Névéol A, Shooshan S, Claveau V, Automatic inference of indexing rules for MEDLINE. BMC Bioinformatics, 2008 **9**:S11.

[22] Gay C, Kayaalp M, Aronson A, Semi-automatic indexing of full text biomedical articles. AMIA Annu Symp Proc, 2005 **2005**:271-5.

[23] Yetisgen-Yildiz M, Pratt W, The effect of feature representation on medline document classification. AMIA Annu Symp Proc, 2005 **2005**:849-53.

[24] Rak R, Kurgan L, Reformat M, Multilabel associative classification categorization of MEDLINE articles into MeSH keywords. IEEE Eng Med Biol Mag, 2007 **26**:47-55.

[25] Jimeno-Yepes A, Mork J, Demner-Fushman D, Aronson A, Comparison and combination of several mesh indexing approaches. AMIA Annu Symp Proc, 2013 **2013**:709-18.

[26] Tsafnat G, Glasziou P, Choong M, Dunn A, Galgani F, Coiera E, Systematic review automation technologies. Syst Rev, 2014 **3**:74.

Address for correspondence

Adam G. Dunn, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 2109, NSW, Australia.
Email: adam.dunn@mq.edu.au

This page intentionally left blank

Posters

This page intentionally left blank

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-875

Characterization of Help Desk issues After the Implementation of an Emergency Department Electronic Health Record

Daniel Capurro[a], Mauricio Soto[b], Patricio Giacaman[a], Silvia Catalán[c]

[a] *Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[b] *Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[c] *Department of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

Abstract

Electronic health records (EHRs) can produce significant disruption when first implemented. Successful implementations depend on the availability of technical and clinical support. We present a description of the frequency and types of issues raised during the first 12 months after the implementation of an EHR at a teaching hospital in Santiago, Chile.

Keywords:
Electronic health records; Emergency department.

Introduction

Electronic health records (EHRs) are information systems that produce significant impacts on usual clinical and administrative workflows, affecting users as well as clinical and administrative systems [1]. Although significant planning and preparation happens before they go live, EHRs produce many issues during the initial go-live and during the weeks and months following the initial date of full operation. This study sought to characterize the frequency and type of issues raised to the local help desk during the first year after the go-live of a new emergency department EHR.

Materials and Methods

This study was conducted in the Emergency Department (ED) of an urban teaching hospital in Santiago, Chile. This ED takes care of an average of 220 patients a day. Before the EHR the department used a paper medical record and paper medical orders; clinicians had access to radiology and laboratory informatin systems. The implemented system included full patient documentation and electronic orders and prescriptions.

The implementation team included experts that provided on-site, 24/7 support to clinical and administrative personnel. These professionals documented and notified issues and created help desk tickets using a smartphone application. Every issue was classified as (a) user training (b) clinical contents (c) vendor issues (d) local Information Technology (IT) team issues (e) Hardware issues and (f) User suggestions. We provide a description of the frequency and type of issues raised during the first 12 months after the ED EHR go live.

Results

The EHR went live on February 13, 2013. During the first 12 months of operation, 3372 tickets were raised by users working at the Emergency Department. As expected, a large proportion of issues were raised during the initial months of operation, with 64% of them raised in the initial 6 months. In addition to decreasing with time, we also observed a change in the distribution of types of issues. The distribution is described in Figure 1.

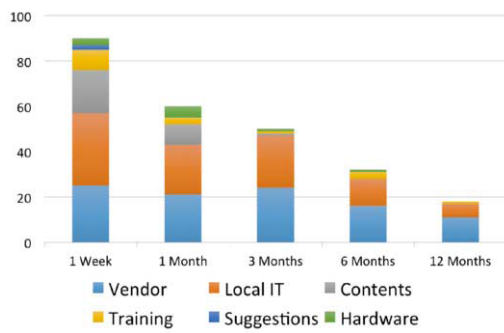


Figure 1- Frequency and distribution of EMR implementation issues (7-day periods at different times after the initial go-live)

All issues decreased their frequency over time but issues related to user training, clinical contents and hardware decreased the most.

Conclusion

As expected, most help desk issues decreased with time; however, issues related to the local IT team—such as internet connectivity, server uptime—and the EHR vendor were still present at 12 months of operation. These findings might suggest that implementation teams should carefully consider the amount of support needed from their vendors and local IT teams.

References

[1] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. JAMIA 2004; 11(2): 104-112.

Address for correspondence

Daniel Capurro, MD, PhD, Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile. Lira 63, Santiago. dcapurro@med.puc.cl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-876

The e-NutriHS: a web-based system for a Brazilian cohort study

Luciana D. Folchetti[1], Isis T. da Silva[1], Bianca de Almeida Pititto[1] and Sandra R. G. Ferreira[1]

[1]*Department of Nutrition, School of Public Health, University of Sao Paulo, SP, Brazil*

Abstract

The e-NutriHS is a web-based system developed to gather online information on health of a cohort of college students and graduates in nutrition. It consists of six validated and internationally recognized questionnaires regarding demographic and socioeconomic data, dietary habits, physical activity level, alcohol and tobacco use, anti-fat attitudes and personal and family histories. Our software and respective database is hosted in the School of Public Health server and is based on free programming languages. An e-NutriHS prototype was created preceding online deployment. An improved version of the website was released based on 20 volunteers' opinions. A total of 503 users were registered. Considering that web-based systems produce reliable data, are easy to use, less costly and are less time-consuming, we conclude that our experience deserves to be shared, particularly with middle income economy countries.

Keywords:
e-Health, Web-based system, Cohort study, Nutrition.

Introduction

In recent years, the use of the Internet, in particular for healthcare information provision, is markedly increasing. In light of this growth, researchers of health sciences have invested in learning how to use the Web for data collection and reporting. Research studies in epidemiology are often conducted with hundreds or thousands participants, followed for long periods of time. The development of web-based online self-administered systems employing validated health-related surveys is much warranted. Such systems to obtain data related to health would be particularly important in developing countries where resources for researches are limited.

Materials and Methods

The Nutritionist Health Study (NutriHS) was conceived in 2013 by our research group from the School of Public Health of the University of São Paulo (USP), Brazil. The NutriHS is a cohort study mainly involving students and some graduates of Nutrition Graduation Courses from universities located in the State of Sao Paulo. For this first phase, launched in 2014, participants are being invited to participate in the study in which health-related data will be collected at a 3-year interval. The ethical committee at the School of Public Health of USP approved this study. A web-based system was developed to gather information online for this specific subset of the Brazilian population, the e-NutriHS.

The e-NutriHS website is based on free programming languages and with the University of São Paulo Informatics Center support, which hosted the website and respective database in the School of Public Health server (http://www.fsp.usp.br/nutrihs/index.html). For its design, the colors were carefully selected and ease-of-use was frequently reviewed until a user-friendly version of the website was obtained. A logotype was created, reinforcing the impact of increasing knowledge about nutrition on health. In addition to registration data, the e-NutriHS includes six validated and internationally recognized questionnaires regarding demographic and socioeconomic data, dietary habits, physical activity level, alcohol and tobacco use, anti-fat attitudes and personal and family histories. The length of each questionnaire page was defined based on a comfortable time for completion.

Preceding online deployment of the website for NuthiHS users, a prototype of e-NutriHS was created and hosted on a temporary url. 20 volunteers and graduates from the School of Public Health aged 25 to 60 years old and who have worked in the health area were invited to test the system and to check its ease-of-use, data entry and website design.

Results

The volunteers considered the website attractive, made favourable comments about the colours used and the interactive elements and suggested no change, after fulfilled the web-based questionnaires. An improved version of website was released based on the volunteers' opinions.

In March 2014, NutriHS participants have started to use the online web-based system. A total of 503 users were registered: 459 college students and 44 graduates. Participants were predominantly female (93%) which was expected since the majority of nutrition students are women. The mean age of the sample was 23.8 (SD 6.6) years; 63.5% were white and 84.6% were single. Preliminary lifestyle data are available, allowing for comparisons of students' profile at the beginning and end of the course. These conditions are in contrast with the nutritional epidemiological studies, which require a long period to get results.

Conclusions

Preliminary findings suggest that the feasibility of e-NutriHS may be useful to obtain data for epidemiological studies, accelerate data gathering, minimize duplicity of information and has potential to produce reliable data and identify health disorders.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-877

Problem Oriented Medical Record: Characterizing the Use of the Problem List at Hospital Italiano de Buenos Aires

Mariano Franco[a], Maria Victoria Giussi Bordoni[a], Carlos Otero[a], Mariana Clara Landoni[a], Sonia Benitez[a], Damian Borbolla[a], Daniel Luna[a]

[a] *Health Informatics Department, Hospital Italiano de Buenos Aires*

Abstract

Problem oriented medical record (POMR) was born in late sixties. Expecting an ordered, complete and updated medical record were some of the goals of its founder. Several healthcare institutions have included problem list into their clinical records but some concerns have been reported. These concerns are in reference to their voluminosity, incompleteness and outdatedness. This study attempts to understand how healthcare professionals are using the problem list at Hospital Italiano de Buenos Aires (HIBA). We believe it is essential to understand the local reality applied to our own applications and cultural instances of documentation. This report is the basis from which several improvements could be made in order to meet the goals of Weed´s proposal.

Keywords:

Problem Oriented Medical Record; Problem List; Electronic Health Records.

Introduction

The Problem List was originally proposed by Lawrence Weed in the sixties as part of his recommendation for a POMR. The arrival of Electronic Health Records (EHRs) has suggested that it is possible to have better problem lists. Healthcare professionals at HIBA have used an electronic POMR since 1998. This paper attempts to have an understanding of the use of the problem list by professionals at HIBA as part of a plan for improvement of this instrument.

Methods

HIBA is a non-profit health care academic center. It has implemented an in-house Healthcare Information System (HIS). The hospital has an insurance plan (Plan de Salud - PS) with 150,000 members. The EHR is a fully-implemented web based, problem oriented, patient centered record with customized functionalities depending on the level of care (outpatient, inpatient, emergency care and home care). We conducted a retrospective cohort e analysis. The cohort included all new problems recorded from January 1st 2012 to December 31st 2013. A secondary de-identified database administered by the area of Biostatistics and Business Intelligence from HIBA´s Department of Health Informatics was consulted for this study. For the analysis of the problem list, the following criterion was agreed upon: all problems in all settings were taken into the sampling.

Results

During the period, there was a total of 2,478,545 new problems. Of these, 42.47 % were associated with at least one medical note. An average of 5.45 problems per patient was calculated. 35.6% of the total sample were created in the outpatient setting, 11.9% in inpatient hospitalization area, and 52.5% in the emergency. Regarding the status of problems, 96.87% were active, 0.70% problems were passive, and 2.12% problems were resolved. Grouping the problems studied by year, and day of charge, we found no significant differences. The twenty most common problems accounted for 30% of all problems. Among these, the 3 most used were fever, health check and malaise. From the top twenty problems, 38% were general symptoms or usual care situations, 23% respiratory diseases, and 39% represented other organs or systems. Of all newly created problems, 73% belonged to patients covered by affiliated insurers, corresponding to 34% of all visits in the period under study. The remaining new problems were generated during visits covered by PS. By analyzing the problems by age, the first segment of 0-5 years old had the largest proportion.

Conclusion

The first step to understand the difficulties is to determine how problem list is used. We conducted an analysis of our POMRs as a basis to generate our own strategies to improve their use, and that goes beyond the implementation of guidelines and institutional policies. Considering the difference in problems creation in the different age groups, we believe that this variability is because the problem list is created in our system once in a lifetime and it is reused for each query. Regarding the most frequent diagnoses, it is not uncommon that most of these instances are due to usual care or general pathologies. In this sense, we expected that the 20 most common problems covered a percentage higher than 30% that we found. We believe this could be due to the large granularity or level of detail that our EHR allows to incorporate. This generates a greater dispersion in the use of problems.

Address for correspondence

Dr. Mariano Franco | Hospital Italiano de Buenos Aires
mariano.franco@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-878

Secretaries' role in EHR Documentation and the Implications of Establishing a Structured EHR System

Monika Alise Johansen[a,b], Åse-Merete Pedersen[a], Gunnar Ellingsen[b]

[a] *Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[b] *Telemedicine and e-Health Research Group, Arctic University of Norway, Tromsø, Norway*

Abstract

Secretaries play an important quality assurance role in today's medical record production. This study aimed to identify quality assurance tasks that a future system cannot easily compensate for when developing a new structured EHR in which the physicians do the writing themselves. The study identified two tasks, which we suggest should also be performed by secretaries in the future.

Keywords:
Electronic health records; structured EHR; dictation

Introduction

Dictation dominates today's medical record production [1]. The introduction of highly structured electronic health records (EHR) implies that physicians have to record the data in the EHR system — that is, write the patient documentation themselves.

Speed, accuracy and completeness improve when doctors use structured templates instead of dictation [2]. Doctors who write patient journals themselves seem to write more briefly and in a way that makes the documentation more readable [3]. A large-scale study in primary care setting showed improved quality of patient care when physicians entered information directly into the structured records [4]. In addition, the documentation is available much earlier - half an hour after the consultation - compared to 374 hours later if the journal note is dictated [5].

However, doctors' change of work practice from dictating to recording may overlook the fact that secretaries play an important quality assurance role in today's EHR production [1]. This study aimed to investigate secretaries' quality assurance role and to identify work that cannot be easily compensated for when designing and establishing a new structured EHR system in which the physicians write the patient documentation themselves.

Methods

Both qualitative and quantitative methods were applied. For the qualitative research, an interpretative study approach was used and data were collected through informal discussions and focus group interviews. In addition, secretaries were asked to fill out one anonymised survey per day, over two consecutive weeks to record their quality assurance work.

Results

The results revealed that secretaries perform numerous quality assurance tasks. Two of the quality assurance tasks were regarded as especially challenging for the future EHR system to compensate for. These work tasks were reported on 28 of the survey forms, representing an equal number of workdays. The first task was to correct the instances in which the same patients had been registered twice or more in the EHR; this issue was corrected 118 times. In addition, the secretaries made 169 corrections due to incorrect patient names or other incorrect essential patient information. This demonstrates that during 28 days of work, each secretary conducted an average of more than 10 very important corrections that could be challenging for the future system to compensate for.

Discussion and Conclusion

Today's secretaries perform numerous important corrections and quality assurance tasks in EHR production. When designing and developing the new structured EHR system, as many as possible of these tasks need to be compensated for by the system. However, to support doctors, we suggest that the following two quality assurance tasks should also be performed by secretaries in the future, as these tasks are very time-consuming and cannot be easily compensated for in the new structured EHR system. These tasks are: correcting the patients' registration and correcting important patient information when mistakes have been identified.

References

[1] Garcia AC, et al. Understanding the work of medical transcriptionists in the production of medical records. Health Informatics J, 2010, vol. 16, no. 2, pp. 87–100.

[2] Cowan DA, et al. Electronic templates versus dictation for the completion of Mohs micrographic surgery operative notes. Dermatol Surg, 2007, vol. 33, no. 5, pp. 588–95.

[3] Vikkelsø S. Subtle redistribution of work, attention and risks: Electronic patient records and organisational consequences. SJIS, 2005, vol. 17, no. 1, pp. 1–30.

[4] Linder JA, et al. Method of electronic health record documentation and quality of primary care. J Am Med Inform Assoc, 2012, vol. 19, no. 6, pp. 1019–24.

[5] Laflamme MR, et al. Efficiency, comprehensiveness and cost-effectiveness when comparing dictation and electronic templates for operative reports. AMIA Annu Symp Proc, 2005, pp. 425–9.

Address for correspondence

Monika A. Johansen, Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, N-9038 Tromsø, Norway; E-mail: monika.johansen@telemed.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-879

Development of Markup Language for Medical Record Charting: A Charting Language

Won-Mo Jung[a], Younbyoung Chae[a], Bo-Hyoung Jang[b]

[a] *Acupuncture and Meridian Science Research Center, College of Korean Medicine, Kyung Hee University*
[b] *Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University*

Abstract

Nowadays a lot of trials for collecting electronic medical records (EMRs) exist. However, structuring data format for EMR is an especially labour-intensive task for practitioners. Here we propose a new mark-up language for medical record charting (called Charting Language), which borrows useful properties from programming languages. Thus, with Charting Language, the text data described in dynamic situation can be easily used to extract information.

Keywords:
Electronic Medical Record; Medical Language Processing.

Introduction

In these days, organizing collection of electronic medical records (EMRs) is regarded as fruitful attempts and many trials have been made in various ways [1]. Most of the platforms being used to collect medical records require practitioners to fill structured data forms. Yet, it might be too strict and rigid form for real clinical situations where dynamic interaction between practitioners and patients exists. Whereas, when the practitioners describe medical records in free-text form, it is difficult to extract data for secondary analysis. Thus, trade-off exists between structured and free-text EMR system. [2] Standardized information model might be one of solutions to the problem, such as HL7 CDA and CCR of ASTM International. The standards are useful references for the process of developing interoperable EMR systems, but the standards do not consider the situation when the individual practitioners construct EMR data directly using the standards [3]. Here, we propose a mark-up language for charting medical record (called Charting Language) by which individual practitioner constructs their EMR structure without constraints and with ease at each treatment.

Methods

The charting language we propose borrows useful properties from programming languages. The charting language is a mark-up language with easy and simple syntax to describe medical records by practitioners in clinical situations. We defined a language grammar that has a hierarchical structure of compound literals (object or section) consisting of primitive literals. The main primitive literal is a medical term that can be imported from a shared library. Because the medical records written by Charting Language has specifically defined syntax and uses medical terms with semantic interoperability, the medical records can be extracted for secondary use of data, like compiling process in the normal programming language. Additionally, within a defined syntax, it provides the capacity to write down medical records without the constraint of the structure within a defined language.

Results

Figure 1 – The abstract syntax and an example.

Syntax of the language was defined in XML format following tmLanguage rule used in popular text editor softwares. For functional convenience, additional plug-ins and snippets were also implemented to help efficient documenting.

References

[1] Davidson E, & Chiasson M. Contextual influences on technology use mediation: a comparative analysis of electronic medical record systems. European Journal of Information Systems 2005: 14(1): 6-18.

[2] Liu B et al. Data processing and analysis in real-world traditional Chinese medicine clinical data: challenges and approaches. Statistics in medicine 2012: 31.7: 653-660.

[3] Ferranti JM et al. The Clinical Document Architecture and the Continuity of Care Record. A Critical Analysis. Journal of the American Medical Informatics Association 2006: 13(3): 245–252.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-880

Using Electronic Medical Record Data to Improve HIV Patient Monitoring, Clinical Decision-Making, and Quality Improvement: Lessons from Rwanda

Nadine Karema[a], Anatole Manzi[a], Erick Gaju[b], Michèle Kayiganwa[b], Eric Remera[c], Alphonse Nshimyiryo[a], Gad Niyibizi[c], Adeline Dukuze[c], Neil Gupta[a, d], Cheryl Amoroso[a]

[a]Partners In Health/Inshuti Mu Buzima, Rwanda
[b]Rwanda Ministry of Health
[c]Rwanda Biomedical Center
[d]Division of Global Health Equity, Brigham and Women's Hospital, Boston, USA

Abstract

In developing countries, clinical guidelines and patient follow-up are primarily paper-based. We describe the use of Electronic Medical Record data for evidence-based clinical decisions and improved HIV patients monitoring in rural Rwanda.

Keywords:
Electronic Medical Records; HIV; Viral load.

Introduction

In 2006 and 2010, two innovative systems were respectively implemented in Rwanda through the joint effort of Partners in Health (PIH) and the Ministry of Health (MOH): the Electronic Medical Record (EMR) system, and the Mentorship, Enhanced Supervision and Quality Improvement Program (MESH-QI). MESH-QI mentors used EMR as a detection and monitoring tool to identify and design QI interventions.

Methods

A QI intervention focused on clinical decisions and HIV patients monitoring was conducted by 2 nurse mentors in 14 health facilities of Kirehe district. The aim was to improve rates of eligible HIV patients tested for viral load (VL) from 2.8% to at least 50% within 3 months (Nov 2012-Feb 2013). The number of each facility patients due for VL testing, and proportions of those with test results were aggregated into EMR reports reviewed and discussed during data sharing meetings held with health facilities leaders. EMR was used to track site performance, and Chi-squared test was used to measure differences pre- and post-intervention.

Results

Figure 1 reports changes in VL testing before intervention [N = 2223 (children = 135 & adult = 2088)] and after intervention [N = 2387 (children = 138 & adult = 2249)].

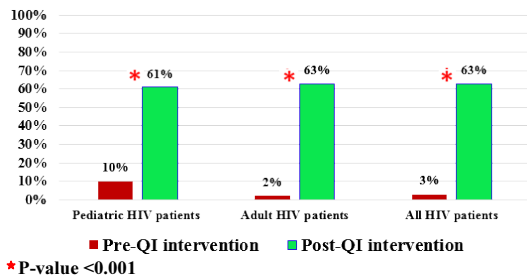


Figure 1- Changes in viral load testing

Conclusion

Use of EMR data may contribute to improve patient monitoring and evidence-based clinical decisions. These findings suggest EMR as a potential source and reference for HIV clinical record keeping, reporting, decision-making, and data-driven QI.

Address for correspondence

For more information: nkarema@pih.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-881

Restructuring an EHR system and the Medical Markup Language (MML) standard to improve interoperability by archetype technology

Shinji Kobayashi[a], Naoto Kume[a], Hiroyuki Yoshihara[b]

[a] *The EHR research unit, Kyoto University, Kyoto, Japan*
[b] *Miyazaki University Hospital, Miyazaki, Japan*

Abstract

In 2001, we developed an EHR system for regional healthcare information inter-exchange and to provide individual patient data to patients. This system was adopted in three regions in Japan. We also developed a Medical Markup Language (MML) standard for inter- and intra-hospital communications.

The system was built on a legacy platform, however, and had not been appropriately maintained or updated to meet clinical requirements. To improve future maintenance costs, we reconstructed the EHR system using archetype technology on the Ruby on Rails platform, and generated MML equivalent forms from archetypes. The system was deployed as a cloud-based system for preliminary use as a regional EHR. The system now has the capability to catch up with new requirements, maintaining semantic interoperability with archetype technology. It is also more flexible than the legacy EHR system.

***Key words**:* archetype, EHR, ISO 13606, MML, openEHR

Introduction

In the Dolphin project, we developed an EHR system for regional health providers to share clinical records and provide individual patient data to patients [1]. We also developed an XML-based standard called Medical Markup Language (MML) to share clinical data within and among hospitals [2]. This EHR system has been working successfully with more than 6000 patients in three regions in Japan. However, this system was not sufficiently maintained or updated because it was built on a legacy web framework, written in the now old-fashioned Perl language. Therefore, we decided to reconstruct the EHR system using common clinical models designed on the ISO 13606 standard with the openEHR specification [3] to improve flexibility.

Methods

We adopted Ruby on Rails as a web framework and a Ruby implementation of the openEHR specification [4]. We redesigned web forms and MML equivalent messages based on clinical concepts involving archetypes in the openEHR clinical knowledge manager (CKM) (http://www.openehr.org/ckm/), with or without modification, plus newly created archetypes. We designed archetypes using the Ocean Archetype Editor and redesigned forms and messages using the Ocean Template Designer.

Results

All of the EHR data components were redesigned using archetypes and are now composed of equivalent templates meeting openEHR specifications. MML messages were generated from archetype components. The web system was implemented for these clinical models with Ruby on Rails and a PostgreSQL database. After functional review, we deployed the system on a cloud platform for preliminary use, for EHR management.

Discussion

Once an EHR system is deployed, it needs maintaining and updating. Moreover, progress in medical knowledge can make systems obsolete. However, archetype technology makes it possible to apply future-proof clinical concepts to construct an EHR system. The restructured EHR system also has improved flexibility for revisions.

Conclusion

EHR systems inevitably need revising. The findings from this work suggest that archetype technology can improve the flexibility of EHR systems.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 24590614.

References

[1] Takada A, Guo s J, Tanaka K, Sato J, Suzuki M, Suenaga T, et al. Dolphin project: cooperative regional clinical system centered on clinical information center. J Med Syst. 2005 Aug;29(4):391–400.

[2] Araki K, Ohashi K, Yamazaki S, Hirose Y, Yamashita Y, Yamamoto R, et al. Medical markup language (MML) for XML-based hospital information interchange. J Med Syst. 2000 Jun;24(3):195–211.

[3] The openEHR Project. [Internet] 1998 [cited 2014 Dec 10]; Available from: http://www.openehr.org/.

[4] Kobayashi S, Tatsukawa A, Ruby implementation of the openEHR specifications, J of Advanced Computational Intelligence and Intelligent Informatics ,2012; 16(1): 42–4.

[5] Kobayashi S, Diego Bosca, Naoto Kume Hiroyuki Yoshihara, Reforming MML(Medical Markup Language) Standard with archetype technology, Indian J of Med Informatics, 2014; 8(2), 57-60

Address for correspondence

Tel: (+81) 75-874-2659, Address: Kyoto research park No 9, 5F, Room 506, Awata-cho 91, Chudoji, Shimogyo-ku, Kyoto, Japan
Email: kobayashi.shinji.5s@kyoto-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-882

Validation for Accuracy of Cancer Diagnosis in Electronic Medical Records Using a Text Mining Method

Yura Lee[a], Soo-Yong Shin[a], Sung-Min Ahn[a,b], Jae-Ho Lee[a,c], Woo-Sung Kim[a,d]

[a]*Department of Biomedical Informatics, Asan Medical Center,*
[b]*Department of Oncology, Asan Medical Center,*
[c]*Department of Emergency Medicine, Asan Medical Center,*
[d]*Department of Pulmonology and Critical Care, Asan Medical Center, Seoul, Republic of Korea*

Abstract

To validate the accuracy of data in electronic medical record, we compared cancer diagnosis and key words in pathologic reports of cancer patients in a tertiary hospital, using text mining method. We investigated in fourteen kinds of cancers that had highest incidence rates in Korea. Approximately two-third (71.0%) of total patients had right match in cancer diagnosis with pathologic report. The ratio of concurrence was the highest (86.3%) in thyroid cancer patients, however, the ratio was the lowest (49.9%) in liver cancer patients. To prevent the errors in data input, a systematic alarm and feedback to clinicians should be required.

Keywords:

Cancer diagnosis; Data quality; Text mining; Data validation

Introduction

Electronic medical record (EMR) system has propagated rapidly, and the size of computerized medical data has been increased exponentially [1]. Accordingly, the quality of medical data has been emphasized not only for patients' safety but also for various purposes. For example, a huge amount of medical data is being managed for supervision of public health. For the convenience and the efficiency of treatment, implementation of inter-center communication via electrical health data is inevitable. In clinical research, the key of confidence of the results is proper size and quality of data.

As an indicator of medical data quality, we checked the diagnosis of cancer patients and compared with the reports of pathologic examinations. We matched the code of diagnosis, recorded based on the ICD-10, with the report of pathologic examinations, via text mining method.

Methods

Fourteen kind of malignant diseases were investigated: thyroid cancer, stomach cancer, colorectal cancer, lung cancer, liver cancer, breast cancer, prostate cancer, pancreatic cancer, bile duct cancer, non-Hodgkin's lymphoma, kidney cancer, cervical cancer, and ovarian cancer. Those cancers had highest prevalence in Korea, reported by National Cancer Information Center of Korea. We included the cancer patients' data from January 1st, 1989 to December 31st, 2013 in Asan Medical Center, Seoul, Korea.

Based on diagnosis codes (ICD-10) typed by clinicians, we categorized the patients into 14 groups of cancers. All pathologic reports of each patient were filtered out by the presence of the specific key words in the text. If there were one or more pathologic reports that contained corresponding key words with the patient's diagnosis, the patient was considered to have correct diagnosis.

Results

During the study period, 236,012 cancer patients had pathologic examination at least once. Approximately two-third of the patients were filtered as there were one or more pathologic reports that contained corresponding key words (167,530/236,012, 71.0%). In thyroid cancer patients, the correspondence of the keyword was the highest among the 14 groups of cancers (86.3%). The lowest proportion ratio of filtration was in the liver cancer group (49.9%), followed by cervical (53.8%) and ovarian cancer (62.2%).

We randomly reviewed the EMR of 100 patients in unfiltered, liver cancer group. Seventy nine of the 100 patients diagnosed liver cancer without pathologic examination. Twelve were non-liver-cancer patients with or without metastasis to liver. Seven patients were missed due to the limitation of filtering though they had proper diagnosis of liver cancer, and the rest two patients had pathologic examination at other hospital.

Conclusion

We could not condemn the one-third unfiltered diagnosis as wrong diagnosis. In liver cancer, the main diagnostic approach is made by the history of underlying liver disease and the results of imaging studies. Therefore, more research should be followed.

However, most of clinicians lack of awareness in correct data input particularly in the cancer patients with vague primary origin or tumors involved in two or more organs.

As clinical decision support systems prevent errors in medical fields, active feedback and alarm system for clinician to guide entering diagnosis may lower the error rate of medical record.

References

[1] F. Martin Sanchez, K. Verspoor: Big data in medicine is driving big changes, IMIA Yearbook 2014.

Address for correspondence

Soo-Yong Shin:

Department of Biomedical Informatics, Asan Medical Center, Seoul, Korea 138-736

Email: sooyoung.shin@amc.seoul.kr

Phone: +82-2-3010-2528

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-883

Developing an Electronic Medical Record for Interlinked Care Services in Haiti

A Marcelin[a], C Perodin[a], C Baja[a], A Bright[a], J Duperval[a], M Duplan[a], F Dérilus[a], S Duda[b], J Pape[a,c]

[a] *Groupe Haïtien d'Etude du Sarcoma de Kaposi et des Infections Opportunistes (GHESKIO),* [b] *Vanderbilt University,* [c] *Weill Medical College of Cornell University*

Abstract

A large clinical care and research organization in Haiti required an electronic medical record system (EMR) to serve the needs of its 30 interlinked clinical programs. After assessing available open source software, the local team designed and implemented a modular proprietary EMR that is improving data quality and patient care. Despite the many benefits of existing open source medical record systems, clinical centers with complex workflow patterns -- even those in resource-limited settings -- should consider developing sustainable, local systems that fit their care model.

Keywords:

Electronic Medical Records; Developing Countries; Haiti.

Introduction

GHESKIO Centers in Port-au-Prince, Haiti, is the largest HIV/AIDS care and treatment institution in the Caribbean, providing continuous medical care since 1982. GHESKIO offers more than 30 separate clinical and support programs, from HIV, tuberculosis, and cholera care to maternal and child health and sexual violence prevention. GHESKIO maintains electronic medical records for over 425,000 patients using a legacy Clipper/dBase system, but the outdated technology could not scale to meet new data collection needs.

Methods

In October 2010, the GHESKIO technical team (all Haitian) engaged all stakeholders to conduct a needs assessment for a new EMR. They identified six core requirements. The system should (1) manage patient information across a wide variety of activities and GHESKIO programs, (2) interface with existing Lab Information and Pharmacy Systems, (3) accommodate a step-wise EMR rollout with minimal disruption to ongoing clinic activities, (4) link GHESKIO's care sites across the city, and (5) allow insertion of temporary data collections to support GHESKIO's research mission. Finally, (6) the EMR must be maintainable locally by GHESKIO developers.

The team installed and evaluated open source EMRs including OpenMRS, OpenClinic, and the Haitian government's iSante EMR [1,2], but the necessary base code changes to fit GHESKIO needs broke the software upgrade path. To meet core requirements, GHESKIO decided to build an EMR. They worked with clinicians to identify key functionality, design data forms, and remove duplicate information.

Results

The GHESKIO EMR was implemented in Microsoft .NET by three full-time programmers over 18 months. The EMR went live in May 2012, with 13 of 16 modules available as of December 2014. Over 425,000 patients records were migrated and clinicians entered data at the point of care and received electronic, rather than paper, notification of lab results, which expedites patient treatment. Medical records were combined for the 1500 patients who had received care at the two GHESKIO facilities. Challenges included data migration, and training of over 90 clinicians in the use of the new EMR. The new EMR is robust in the wake of natural disasters, because data are backed up in different buildings within sites and also between sites. Data comprehensiveness improved with a 100% increase in available data fields. Providers now can code 3,959 symptoms and diagnoses mapped to ICD-10 codes and presented in user-friendly auto-complete fields, compared to the 115 pre-defined symptoms and diagnoses in the old EMR. The EMR has also facilitated the evaluation and publication of outcomes data [3-5].

Conclusion

Existing EMRs designed for resource-limited settings may not meet the needs of highly interlinked care centers with large research programs. GHESKIO's tailored EMR permits the organization to document patient care with greater precision, participate in a broader set of research studies, and improve continuity of care. This report demonstrates the feasibility of locally implementing and maintaining customized EMRs at major clinical care centers in resource-limited settings.

Acknowledgments

This project was funded by NIH, Global Fund and PEPFAR NIAID/NIH U01AI069923.

References

[1] Aminpour F et al. Utilization of open source EHR around the world: A systematic review. J Res Med Sci Off J Med Sci 2014;19:57–64.

[2] Matheson AI et al. Implementation and expansion of an EMR for HIV care and treatment in Haiti: an assessment of system use and the impact of large-scale disruptions. Int J Med Inf 2012;81:244–56.

[3] Koenig SP et al. Cost-Effectiveness of Early Versus Standard ART in HIV-Infected Adults in Haiti. ***PLoS Med*** 2013 8(9)

[4] Noel E et al Attrition from HIV Testing to ART Initiation among Patients Newly Diagnosed with HIV in Haiti. ***JAIDS*** 2013; 62;61-69

[5] Kimmel AD et al. Lives saved by expanding HIV treatment availability in resource-limited settings:example of Haiti. *JAIDS* 2013;63(2):e40-8.

Address for Correspondence:

Adias Marcelin, amarcelin@gheskio.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-884

Applicability of different types of Patient Records for Patient Recruitment Systems

Björn Schreiweis[a], Björn Bergh[a]

[a] *Center for Information Technology and Medical Engineering, University Hospital Heidelberg, Germany*

Abstract

Patient records – types of Electronic Medical Records – are implemented to support patient recruitment. Different types of patient records have not yet been analyzed as to the number of Patient Recruitment System requirements can be found in each type of patient record. According to our analysis, personal electronic health records (PEHRs) tend to allow for most requirements to be found.

Keywords:

Clinical trials as topic, patient recruitment system, PRS, EMR, EHR, PEHR.

Introduction

The increasing availability of electronic health information has resulted in an increasing number of options for the secondary use of health information. One important secondary use is supporting patient recruitment for clinical trials. Clinical trials often fail to complete recruitment on time, inadequately obtain the correct number of recruits and/or experience budgetary problems [1]. Patient Recruitment Systems (PRSs) are used to improve patient recruitment [2]. PRSs are often integrated with Electronic Medical Records (EMRs), but have not been analyzed as to how other types of records match with the requirements of PRSs.

This work analyzes the applicability of different types of patient records for their integration with PRSs by analyzing the requirement implementation capacity of PRSs.

Methods

A requirements analysis resulting in 13 requirements of PRSs was performed; by conducting unstructured interviews and a literature review [3]. A literature search on types of patient records currently implemented was performed and resulting patient records were analyzed as to their different functionalities and operational paradigms. We found several types: EMRs as records implemented within a healthcare institution, Electronic Health Records (EHRs) as records implemented in healthcare networks (cross-institutional), EHRs for a distinct medical condition (ger: Elektronische Fallakte (EFA)), Personal Health Records (PHRs) for patients' documentation on their own medical history and Personal Electronic Health Records (PEHRs) integrating both PHR and EHR capabilities; allowing for the integration of clinical documentation from the EHR, for patients adding documentation on their own (PHR) and also for patients to control the privileges of their PEHR.

Results

We found differences in the capacity to integrate PRS with EMRs, EHRs, EFAs, PHRs and PEHRs.

Integrating PHRs with EMRs suggests 61.5% (8 of 13) of the requirements are implementable. For EHRs, PHR integration suggests 53.8% (7 of 13) of the requirements are implementable. In an EFA, 30.8% (4 of 13) of the required functionalities can be realized by integrating a PRS. Integrating a PRS with PHRs allows for 38.5% (5 of 13) of the requirements to be implemented. PEHRs are fully implementable for 84.6% (11 of 13) of the required functions. Two functions of PRSs are partly implementable in PEHRs, as patient consent can prevent physicians to be informed about contacting eligible patients because patients can choose not to inform the physician that they are eligible.

Discussion

The integration of a PRS within a PEHR environment allows for the comparison between other types of patient records on the topic of their implementation of requirements for PRSs. Data integration and use of medical information for research purposes (such as matching eligibility criteria) are fully controlled by the patient through consent management. A PRS integrated with a PEHR would be a cross-enterprise PRS, a XPRS.

Conclusion

Further research on PRS integration with PEHRs will be performed and intends to lead to architectures capable of informing how the integration of PRS with PEHRs can be achieved successfully.

References

[1] McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, Elbourne DR, Francis D, Garcia J, Roberts I and Snowdon C. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 7(2006): 9.

[2] Cuggia M, Besana P and Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. International journal of medical informatics 80(2011): 371-88.

[3] Schreiweis B and Bergh B. Requirements for a Patient Recruitment System. Stud Health Technol Inform. Accepted for publication 30.01.2015.

Address for correspondence

Björn Schreiweis, Center for Information Technology and Medical Engineering, University Hospital Heidelberg, Speyerer Straße 4, 69115 Heidelberg, mailto:bjoern.schreiweis@med.uni-heidelberg.de, Phone: +49 6221 56 36148

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-885

Evaluating the data completeness in the Electronic Health Record after the Implementation of an Outpatient Electronic Health Record

Mauricio Soto[a], Daniel Capurro[b], Silvia Catalán[c]

[a] *Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[b] *Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[c] *Department of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

Abstract

Electronic health records (EHRs) present an opportunity for quality improvement in health organitations, particularly at the primary health level. However, EHR implementation impacts clinical workflows, and physicians frequently prefer to document in a non-structured way, which ultimately hinders the ability to measure quality indicators. We present an assessment of data completeness—a key data quality indicator—during the first 12 months after the implementation of an EHR at a teaching outpatient center in Santiago, Chile.

Keywords:
Electronic health records, Outpatient, Data completeness.

Introduction

There is a significant gap between the data that are captured during routine clinical care and the structured data needed for secondary analyses [1]. In an attempt to assess the impact of our implementation strategy, this project sought to characterize the level of data completeness—according to our local documentation quality standard—and level of completeness of the different functionalities of the system during the first year after the go-live of the outpatient EHR.

Materials and Methods

This study was conducted in an outpatient clinic, affiliated to an urban teaching hospital in Santiago, Chile with an average of 3000 visits/month. Our implementation included strategies to increase adoption: personalization of documentation using templates, online and classroom training, on-site personal support and the use of audit and feedback based on the level of completeness of EHR use by physicians. Our EHR completeness quality standard involves the proportion of medical visits with complete documentation in all five priority areas: chief complaint, history of present illness, physical examination, clinical diagnosis, and patient suggested plan.

Results

The EHR went live on December 2013. During the first months 49.7% of all clinical encounters met the data completeness quality standard. Figure 1 describes temporal evolution during the 12 months after the implementation. Observed compliance was higher than reported by the literature. Table 1 describes completeness of each area of clinical documentation.

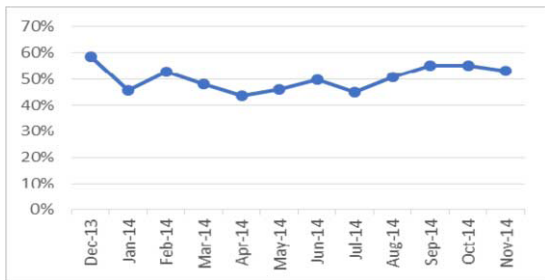


Figure 1- Quality standard compliance of data completeness during 1 year after the EHR implementation

Table 1.

Clinical Documentation Domain	Completeness	Roth 2013[2]
Chief complaint	91%	12%
History of present illness	71%	30%
Physical exam	76%	14%
Diagnosis	100%	72%
Suggested plan	73%	

Conclusions

After one year, we have observed high and stable rates of data completeness. This suggests successful adoption strategies implemented before and after implementing the new EHR.

References

[1] Chan KS, Fowles JB and Weiner JP. Electronic Health Records and the reliability and Validity of Quality Measures: A Review of the Literature. Med Care Res Rev 2010; 67(5): 503-527.

[2] Roth CP, Lim YW, Pevnick JM, Asch SM and McGlynn EA. The Challenge of Measuring Quality of Care From the Electronic Health Record. Am J Med Qual 2009; 24; 385-394.

Address for correspondence

Mauricio Soto, MD, Ms Sc, Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile. Vicuña Mackena 4686, Santiago. msotod@med.puc.cl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-886

Health Information Needs for Child-in-Care

Cori Thompson[a]

[a] *School of Health Information Science, University of Victoria, BC, Canada*

Abstract

When a child is taken into care the State is legally required to provide healthcare. Six forms were reviewed by medical care providers, foster parents and government social workers to understand their health information needs in caring and planning for child-in-care. The qualitative study used a sociotechnical systems framework and NVivo 10 for encoding. Interview findings include that the forms, if completed and available, meet most basic health information needs with additional forms used for complex health needs. The majority of participants indicate interest in electronic options. Focus groups will enable further study.

Keywords:

Foster Care; Health Record

Introduction

Children are considered vulnerable populations at risk for not having health needs met [1]. The State or Province has a duty to provide healthcare to a child- or youth-in-care (CIC) [2]. Health records tailored to CIC exist [3]. A study was conducted with medical care providers, foster parents, and social workers from a provincial social sector ministry that provides protection services to children and families. The study asked what health information the participants need to provide care to and plan for child-in-care.

Methods

Qualitative research with a sociotechnical systems (STS) framework [4] was used to understand the findings. The STS concepts include: social and technical analysis; variance control; linear / non linear tasks; blurring of roles; information sharing; coalitions; and deliberations and decisions [4]. The 19-question interview tool included reviewing six forms: newborn record, child medical record (CMR), birth family medical and social history, care plan, Looking After Children health section, and healthcare passport. The CMR was the main form studied. It is a snapshot of the overall health of the child when a child transitions (eg. into care). Participants were asked for professional opinions. Interviews were recorded, transcribed and are inductively encoded using NVivo 10 QSR.

Results

From November 2013 to 2014 thirty-three interviews were held with thirty-nine participants in six urban, rural and/or remote communities. Most interviews were conducted on site. Findings suggest the forms, when completed and available, generally work, with additional forms needed for complex care needs. The CIC health record information starts with the birth family social history and history of illness and the birth mother's pregnancy up to the CIC's health and development at age 19, including dental care and placements. At age 19 CICs have access to their information. Reunification or permanency planning goal impacts the information collection. Primary data sources, temporal dependency and some redundancy were identified between forms. Medical care providers, foster parents and government social workers may each change while the CIC is in care. The CIC may also transition in and out of care. The legal guardian role, which consents to treatment and information sharing, may transfer between biological parents and ministry. All these changes increase the system's complexity and may increase risk of record fragmentation. Missing information might include: biological parent's health information, CIC's childhood illnesses, records not gathered prior to assessment. Good communication, help with timely record collection, education, healthcare advocacy, clearer guidelines on information sharing, and funding were identified as potential enablers. The majority of participants indicate an interest in electronic options.

Conclusion

If completed and available, the six forms seem to generally meet basic information needs, with additional forms needed for complex health needs. Careful consideration needs to be given to any electronic option as technology adds complexity [5] to an already complex system. Study limitations include that one researcher conducted, transcribed, and encoded the interviews. Focus groups will enable further study.

Acknowledgments

A sincere thank you to the participants and to the individuals and organizations guiding this study.

References

[1] Kliegman R, Nelson WE. Nelson textbook of pediatrics. 19th ed. Philadelphia: Elsevier/Saunders, 2011.

[2] United Nations Human Rights. Convention on the Rights of the Child. 1989.

[3] Thompson C, Lau FY. A scoping review on health records for child-in-care. Stud Health Technol Inform 2013: 183: 43-8.

[4] Pava C. Redesigning sociotechnical systems design: Concepts and methods for the 1990s. The J. Of Appl. Behavioral Science 1986: 22(3): 201-221.

[5] Drummond WH, Ferranti JM, Lehmann CU, Lighter DE. Complexity in health information technology systems. In: Lehmann CU, Kim GR and Johnson KB, eds. Pediatrics Informatics: Computer Applications in Child Health: Springer, 2009; pp. 83-117.

Address for correspondence

Corresponding author: cdthomps@uvic.ca

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-887

Video Conferencing Services in Healthcare: One Communication Platform to Support All

Andrius Budrionis[a], Per Hasvold[a], Gunnar Hartvigsen[a], Johan Gustav Bellika[a]

[a]*Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromso, Norway*

Abstract

We present a novel approach to the design of video conferencing (VC) systems, taking advantage of recent technological achievements in web-based implementation. Delivering VC functionality as a service over the Internet opens new grounds for easier integration, support, and application in many scenarios, since hardware-agnostic ad-hoc VC connections are a feature of the proposed architecture. Validity is demonstrated through latency measures in a surgical telementoring service and comparing them to reported thresholds.

Keywords:

Telementoring; Scalable; WebRTC; Latency.

Introduction

Video conferencing (VC) is a common core technology in telemedical systems. The integration of VC into clinical workflows is associated with high complexity, which prevents exploitation of the potential in various scenarios. Platform dependency, client software installation and updates, complications in traversing complex network topologies, and firewalls are common issues which often need to be taken care of by users. The complexity of employing such systems in clinical settings, while ensuring high availability and managing maintenance costs is just one reason for the slow progress and adoption of the approach, regardless of its high potential.

The use of VC is still limited to planned meetings that are configured and tested in advance. Even the most user-friendly VC systems – like Skype, Google Hangouts, and more advanced solutions by Cisco or Polycom – require registration, software downloads, and installation. However, numerous use cases call for a more flexible and user-friendly VC interface. For instance, emergency surgical telementoring sessions, or GP-patient consultation over VC. Devices used remotely cannot be set up in advance and need to work on-the-fly.

Materials and Methods

Web Real-Time Communication (webRTC) was selected as a backbone for implementing a general purpose VC infrastructure meeting the needs of the evolving healthcare environment [1]. It aims to provide a customizable VC gateway for developing platform-agnostic clinical applications featuring low usability threshold ad-hoc connections between peers.

One-way latency of the proposed system was measured to assess compliance with the reported thresholds in surgical telementoring [2]. VC node was started on an i5 Macbook Pro (late 2013) with 8GB of RAM, and accessed from a 10.1" Asus MeMO tablet computer with 1 GB of RAM, running Android 4.2, on a dedicated network. Two light sensors connected to an Arduino circuit board registered lighting changes on the screens of the interacting peers when a white sheet of paper was put in front of the camera. The experiment was repeated under different network loads.

Results

The generic architecture in Figure 1 provided a framework for developing platform-agnostic VC services. One-way video latency measures averaged at 226.7 ms under perfect network conditions, and 325.7 and 338.7 under 0.5 MB/s and 1 MB/s data traffic respectively. The use of the service added a 133.2 ms delay to the native latency of the network.

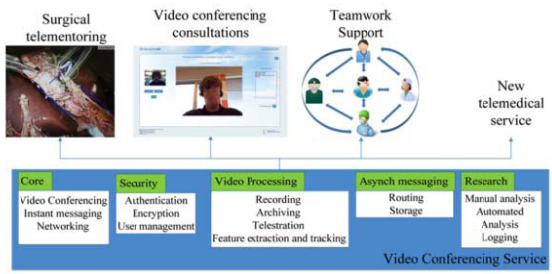


Figure 1. Generic architecture of a VC service

Conclusion

We present a generic architecture providing a secure, highly available, and easy to integrate VC channel that adds no additional maintenance and support costs to hospitals or patients. No matter what technical infrastructure lies between the two interacting parties, the only requirements for establishing the VC session is an updated web browser and an Internet connection. The feasibility of using the architecture to build clinical services was demonstrated and revealed promising results; one-way video transmission latency was acceptable in a surgical telementoring scenario. Assuming lower latency requirements in less safety-critical use cases enables easy adoption of telemedical practices in a wide spectrum of scenarios within both primary and secondary care, which are currently not supported by existing systems.

References

[1] "WebRTC 1.0: Real-time Communication Between Browsers. W3C Editor's Draft 27 January 2014."

[2] M. Anvari, T. Broderick, H. Stein, T. Chapman, M. Ghodoussi, D. W. Birch, C. Mckinley, P. Trudeau, S. Dutta, and C. H. Goldsmith, "The impact of latency on surgical precision and task completion during robotic-assisted remote telepresence surgery," *Comput. Aided Surg.*, vol. 10, no. 2, pp. 93–99, Jan. 2005.

Address for correspondence

Andrius.Budrionis@telemed.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-888

Implementation of a Teleconsultation Service in the Primary Health Care in Brazil

Rodrigo V. Andreão[a], Marcelo Q. Schimidt[b], Thiago D. Sarti[b], Solange L. Corradi[c]

[a] *Federal Institut of Espirito Santo, Vitoria, ES, Brazil*
[b] *Federal University of Espirito Santo, Vitoria, ES, Brazil*
[c] *Planning Department, Espirito Santo Secretary of Health, ES, Brazil*

Abstract

The implementation of a Teleconsultation service in the primary health care is a great challenge. This work presents the effort made in the Espirito Santo Telehealth Project to consolidate the teleconsultation service. The impact of each strategy made was evaluated in terms of the number of teleconsultations produced. The combination of periodic visits and the promotion of local workshops were responsible for a sustainable use of the service for the last six months.

Keywords:
Telehealth; Teleconsultation; Primary Health Care.

Introduction

Considering the great contrast in offering the primary health care (PHC) in Brazil, the Brazilian Telehealth Program in Espirito Santo state, is being implemented aiming at providing the increase of the resolution of most clinical cases and the reduction of the forwarding of patients at the PHC level.

The main challenge of the project is to introduce this new program in the health professional routine, covering 78 cities of the Espirito Santo state, more than two hundred of health units and thousands of health professionals.

There are many experiences in Brazil and in other countries which have approached such subject. However, our ongoing experience shows that new strategies must be taken into account in order to place the telehealth services in the routine of the health professionals.

In this work, we present our strategy during the implementation of the Telehealth Program in Espirito Santo, focusing on the Teleconsultation service. Our results, which were based on a survey answered by the health professionals, confirm the fact that there is a demand on the teleconsultation service.

Materials and Methods

The teleconsultation service consists in offering a qualified second opinion in response to the questions made by local health care provider. This second opinion is a text message made by family physicians, nurse practitioners, and others specialists with expertise in telehealth and general practice and is based on the best available scientific evidence. This kind of collaboration among health professionals takes place in a web based plateform called Salus, where the access is secure and controlled by a personal password.

The service is offered to a health unit city since it meets the following requirements: a. health secretary approval; b. health unit providing a computer connected to the internet; and c. health professionals must follow a short course on the Telehealth project and the Salus platform.

However, during the implementation of our teleconsultation service, we noticed that the requirements above were not enough to the success of the project. In fact, placing the teleconsultation service in the routine of the health professionals remains a great challenge. In order to face this, we included the following strategies: i. phone calls to the manager of the health unit and to the doctor; ii. periodic visits to the health units connected to the project; iii. local workshops exchanging experiences among health professionals; and iv. a financial support to the cities which is participating to the telehealth project.

Results and Discussion

The implementation of the teleconsultation service started on February 2012, when the first health unit was connected. We noticed that during one year and a half, the teleconsultation service did not progress enough and the number of teleconsulations did not cross 29 teleconsultations per month, produced by 19 health centers. As a consequence, we decided to telephone the managers and doctors of the PHC. After that, we detected a sudden increase of the teleconsultations up to 62 per month, produced by 32 health centers. However, the phone calls did not produce a sustainable use because the number of teleconsultations decreased one month later. In order to raise the awareness towards the teleconsulting, we promoted three other strategies, particularly local workshops exchaging experiences about the benefits of the teleconsultings and periodic visits to the health centers of telehealth. These strategies produced a sharp increase up to 124 teleconsultings/month, produced by 30 health centers, which remained sustainable for six months confirming that there is a demand to be met. We believe that repeating these strategies every year the teleconsultation service can definitly became as a part of the routine of the health professionals.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-889

Introducing Home Blood Pressure Telemonitoring for Children with Hypertension

McKenzie Bedra, MPH, Joseph Finkelstein, MD, PhD

Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

The goal of this study was to introduce home blood pressure (BP) telemonitoring in children with hypertension and to assess the feasibility of this approach. Acceptance of the system was assessed by attitudinal survey and semi-structured qualitative interview. Qualitative interview results showed consistently positive comments for content, interface and process components. BP measurements obtained by self-testing were as reliable as Dinamap measurements. The home telemonitoring system was positively accepted, easy to use and found to be helpful by participants. Home-based BP telemonitoring has significant potential to improve patient-centered delivery in children with hypertension.

Keywords:

Telemonitoring, Hypertension, Children, Disease Management, Self-care

Introduction

The long-term goal of this project is to improve adherence to monitoring of blood pressure and anti-hypertensive medication, thereby decreasing hypertension-related morbidity and mortality in adulthood. To achieve this goal, this study had the following specific aims: 1) develop and evaluate patient/caretaker satisfaction with a telemonitoring system for blood pressure (BP) in children with essential hypertension, which will involve self-monitoring at home and allow prompt reciprocal exchange of medication adherence and BP measurement information between patients and health care providers; 2) evaluate the validity of BP measurements obtained during home telemonitoring by comparing results to those obtained by a trained professional; 3) determine patient adherence to BP monitoring and hypertensive medication after utilizing the home telemonitoring system in comparison to self-reported adherence prior to using the system.

Materials and Methods

The hypertension home telemonitoring system included a portable oscillometric blood pressure monitor and appropriate-sized blood pressure cuff, which was connected via a serial port to a laptop computer. The laptop allowed transmission of blood pressure and patient response data (reporting of symptoms of hypertension or hypotension, and side effects of anti-hypertensive medications) via wireless network to the medical center's secure clinical information system twice a day. The laptop computer's functions were limited so that only home blood pressure monitoring could be performed, analogous to an automated teller machine.

Results

Eight children with essential hypertension were enrolled from a pediatric nephrology clinic. The average amount of time that participants took to complete all components of the home telemonitoring system was 256 ± 125 seconds. The blood pressure symptom diary took an average of 31 ± 20 seconds, while assessment of medication side effects took an average of 33 ± 11 seconds. Self-monitoring of blood pressure took an average of 149 ± 79 seconds and completion of the hypertension education module took 43 ± 26 seconds on average. Overall, there were no significant differences noted in the measurements, however self-monitoring took approximately 55 seconds longer on average as compared to measurement with Dinamap and this difference was significant. Attitudinal survey results demonstrated that 88% of participants reported that working with the computer was not difficult, and 100% reported that the self-testing procedures, blood pressure monitor, symptom diary and medication side effect questions were not complicated. Seventy-five percent felt the self-testing procedure took little time and did not interfere with their usual activities. Sixty-three percent felt slightly to significantly safer while monitored by the home telemonitoring system and felt it was important to know that the self-testing results could be immediately reviewed in the medical center after the test. Overall, 88% would like to use the home telemonitoring system in the future. The following were typical responses that focused on content: "The symptoms that were listed in the symptom diary were helpful. They are symptoms that I actually feel," and, "I thought the education was the best part because I can learn more about high blood pressure and what symptoms I might have." Examples of responses that focused on the interface included: "I thought the home BP machine was easy to use, and the instructions were very clear."

Discussion

Home-based BP telemonitoring has significant potential to improve patient-centered delivery in children with hypertension. Patients and caregivers provided valuable feedback which should be addressed in future generations of the system. Next steps should include definitive evaluation using randomized study design to evaluate impact of home-based telemonitoring on clinical outcomes in children with hypertension.

Acknowledgments

We are very thankful to Dr. Cozumel Pruette for providing access to the study subjects and useful comments on clinical aspects of the study protocol.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-890

Providers Expectations on Telemedicine: A Qualitative Research in a Large Healthcare Network of Latin America

Giussi Bordoni MV, Plazzotta F, Sommer J, Benítez S, García G, Luna D, González B De Quirós F

Health Informatics Department, Hospital Italiano de Buenos Aires

Abstract

The benefits of Telemedicine make it a viable, reliable and useful discipline for dispensing health care. This qualitative study is aimed to understand the expectations, opinions and previous knowledge of the professionals about telemedicine at the Hospital Italiano de Buenos Aires. Results: Professionals realize that Telemedicine is inserted into their usual practice in an informal way. They consider telemedicine as an alternative to the traditional delivery of health care, but are afraid of their role in health care is undermined. Professionals point out very specific applications of Telemedicine such as monitoring the health of patients remotely, drug doses adjustments and sharing clinical information. Conclusion: Results suggest that professionals are not familiar with telemedicine and will be necessary to develop a training plan before implementation.

Keywords:
Telemedicine, Telehealth, Qualitative Research.

Introduction

Healthcare system of Argentina is highly segmented, inequitable and heterogeneous. Three important challenges are, get better access to services, increase the quality of care, and lower the costs. Healthcare Information Technologies comprises a set of tools that offer an opportunity to overcome the challenges and Telemedicine is postulated as a viable, reliable and useful discipline. Several studies have attempted to collect data about healthcare professionals' expectations and opinions on telemedicine. Evidence found is controversial, although most of them report that available technologies are acceptable and easy to integrate, they believe its necessary to establish improvements to make them more reliable and adaptable to users' workflow.

Hospital Italiano de Buenos Aires is planning to develop a new healthcare model based on telemedicine using the Electronic Health Record and the Personal Health record. The aim of this study is to understand healthcare professionals' expectations, opinions and prior knowledge about telemedicine as part of the project's first stage.

Methods

A qualitative study based on in-depth interviews and focus groups was conducted in November and December of 2014. Content analysis was performed using Grounded Theory and NVIVO.

Results

Twenty five professionals (medical and non-medical) participated in 5 in-depth interviews and 2 focus groups. Participants' specialties were Family Medicine, Pediatrics, Psychiatry, Psychology, Nursing, Internal Medicine, Surgery, Dermatology, Nutrition and Lactation. Participants agreed that "Telemedicine" is a form of remote communication:

"I guess its medicine in the world of telecommunications where television and Internet are essential in everyone's life".

Professionals identified phone, email and Skype as tools used to communicate remotely with colleagues or patients:

"I had remote-live-supervisions using Skype, since the other therapist lived far away".

"Sometimes I give medical counseling to patients using Skype, I find it very useful".

Different opinions were emerging referring to the physician-patient relationship, the role of the institution, and the perception they have about this healthcare modality. They report that it's important to have a previous relationship with the patient before using telemedicine:

"It´s useful with patients that I already know because a previous relationship has been established".

Professionals are concerned about the use that patients make of these tools and are afraid of losing professionalism in the doctor-patient relationship:

"Patients use the email to medical consultations and for request to an appointment. Others use "Whatsapp" and this is not formal".

Professionals are also concerned about the role of the institution regarding the regulation of these new practices:

"Patients have no limits: call you late at night for small queries and not for actual emergencies. The institution has the role to create norms and procedures on how to handle these communication devices".

At the chance of using Telemedicine at their daily workflow they foresee communication with other professionals or education contexts. They said Telemedicine is not useful if the reason for consultation required physical examination:

"I think it's a useful tool for people who live far away, but if you have to touch the patient is not helpful".

Conclusions

Results suggest that professionals were not familiar with this kind of healthcare delivery. They found difficult to imagine how telemedicine would be implemented into their daily workflow. For future implementation, it will be necessary to design a training plan for health care professionals and change management strategies.

Address for correspondence

Dr. Giussi Bordoni María Victoria
maria.giussi@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-891

Future Telehealth and Telecare Reference Design based on IoT Technologies: From Remote Monitoring to Smart Collaborative Services with Decision Support

Martin Gerdes, Frank Reichert, Jan Pettersen Nytun, Rune Fensli[a]

[a] *Department of Information and Communication Technology, University of Agder, Norway*

Abstract

The demographic changes are producing aging societies across the world, resulting in greater demands on the health and care systems due to age-related disabilities and chronic diseases. Efficient telehealth and telecare services are needed to control the corresponding expenditures, by supporting increased collaboration between different professional and involving informal health care providers, and by empowering the patients to manage their health and well-being.

Emerging trial systems for remote patient monitoring present preliminary solutions not exempt of certain limitations. We propose a future eHealth reference system architecture and core components, aiming at secure, smarter and more collaborative telehealth and telecare services. The implicit cooperation between the so-far separated domains of consumer well-being services and public telehealth and telecare services will be beneficial for all parties.

Keywords:

eHealth; health informatics; telehealth; telecare; Internet of Things; ontologies; EHR; distributed decision support.

Introduction

An increased efficiency of the development, operation and utilization of telehealth solutions requires a smarter and adaptive device platform for data collection and communication in the point-of-care, distributed reasoning and decision support, and a secure storage and provisioning infrastructure. This has to support the cooperative access to health data and information (Core-EHR, P-EHR, other records) for professional and informal health and care providers, enforcing also privacy protection and access control.

Methods

For the EU FP7 project United4Health a trial system for remote monitoring of COPD patients at home has been developed and deployed in the secured Norwegian National Health Network (NHN). The solution provides routine measurements and questionnaires for the patient, and a secure and collaborative provisioning of the results to professional health care sources via an Information Integration Platform (IIP) and a dedicated NHN telehealth Web portal service.

Results

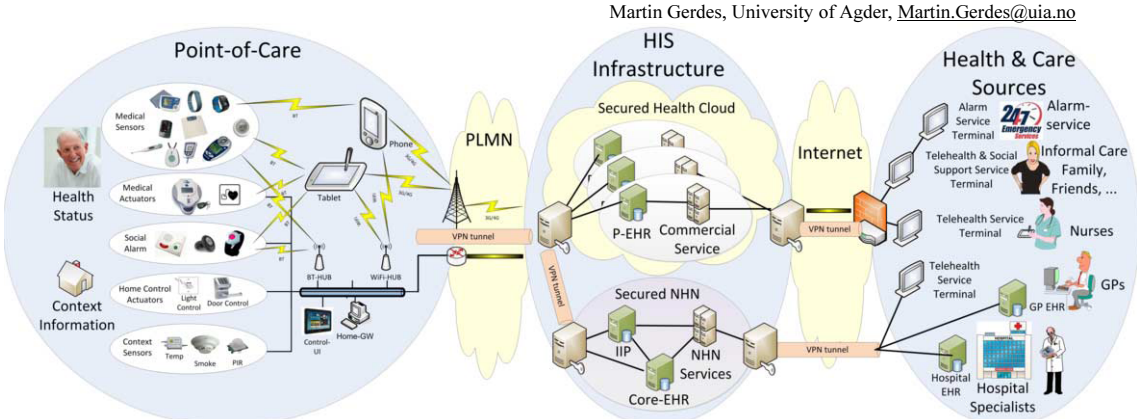
The result is a Secured Health Cloud infrastructure for commercial well-being, health and care services, complementing the NHN infrastructure on top of a common authentication and access control system as part of the overall Health Information System (HIS) infrastructure. The proposed point-of-care device platform will be personalizable for the patients' needs, utilizing smart sensors and ontologies for the individual data and as basis for adaptable, remotely manageable software modules for distributed decision support and user interaction.

Conclusions

The proposed reference architecture provides the basis for secure collaboration between commercial, professional and informal health and care sources, enabling personalized, smarter, and more efficient health and care services.

Address for correspondence

Martin Gerdes, University of Agder, Martin.Gerdes@uia.no



MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-892

The EU-project United4Health: User-Centred Design and Evaluation of a Collaborative Information System for a Norwegian Telehealth Service

Berglind Smaradottir[a], Martin Gerdes[a], Santiago Martinez[a], Rune Fensli[a]

[a] *Centre for eHealth and Healthcare Technology, University of Agder, Norway*

Abstract

This study presents the user-centred design and evaluation process of a Collaborative Information System (CIS), developed for a new telehealth service for remote monitoring of chronic obstructive pulmonary disease patients after hospital discharge. The CIS was designed based on the information gathered in a workshop, where target end-users described the context of use, a telehealth workflow and their preferred ways of interaction with the solution. Evaluation of the iterative refinements were made through user tests, semi-structured interviews and a questionnaire. A field trial reported results on the ease of use and user satisfaction during the interaction with the fully developed system. The implemented CIS was successfully deployed within the secured Norwegian Health Network. The research was a result of cooperation between international partners within the EU FP7 project United4Health.

Keywords:

Health Information Systems; Information and Knowlegde Sharing; Usability; Telehealth; User-Centred Design.

Introduction

The research project United4Health (U4H) [1] aims to evaluate patient experiences with telehealth in Europe. In the Norwegian contribution to U4H, a solution was developed for remote monitoring of chronic obstructive pulmonary disease (COPD) patients. A Collaborative Information System (CIS) was built for information management and efficient communication in a new telehealth service, with a focus on sustainable operation and implementation within the secured Norwegian Health Network. This poster presents a summary of the results from the user-centred design (UCD) and evaluation process of the CIS from operational and qualitative usability perspectives.

Methods

The UCD process started with a workshop where representatives of the different end-user groups described the context of use and the preferred ways of interaction with the solution. The outcome was used to define user requirements to inform the CIS design. Evaluations of iterative refinements which included semi-structured interviews and a Satisfaction Usability Scale (SUS) questionnaire were made with end-users in the usability laboratory of the University of Agder, Norway. A final evaluation of the CIS was performed through a field trial before its deployment. Qualitative methods for data collection and analysis were employed during the UCD process, which had a total duration of 6 months.

Results

Based on the design input from the workshop, the first version of the CIS was implemented. Representative end-users iteratively tested the implementation in the usability laboratory. In the first evaluation, 9 usability issues were identified related to the GUI, among which 1 major issue was related to the colour scheme selected that could interfere with the coloured evaluation of patient *triage*. Consequently, a grey scale colour scheme was advised. Users made 9 suggestions about the functionality of the system, such as having the option to manually override the automatically calculated patient *triage*. The second user evaluation showed that the new GUI colours improved information overview and the manual override functionality was evaluated as successful. Answers from the SUS questionnaire presented high score of satisfaction. Results from the field trial showed that users' overall rating concerning the interactions with the CIS was satisfactory. In addition, the CIS was reported by users as "*easy to navigate*" and the display of critical information at the top of the GUI was valued as "*positive*". One of the participants stated during interview: "*The CIS seems to work well and gives a good overview, most of it is self-explained*".

Conclusion

The UCD process effectively included users' needs in the design process of the Collaborative Information System for a Norwegian telehealth service for remote monitoring of COPD patients after hospital discharge. The workshop with representative end-users allowed gathering of key information regarding data visualisation and system requirements. The evaluation comprised a series of iterative user tests that led to the refinement of the GUI and functionality of the system. In addition, a field trial provided both real-time evaluation and continuous observation of long-term use of a fully developed telehealth system in working environment, contributing to the ecological validity of the study. The CIS has been deployed by the EU FP7 project U4H partners currently being used by three telehealth service centres in Norway.

References

[1] United4Health. European commission- ICT policy support programme and competitiveness and innovation framework programme; [cited 2015 April 10]. Available at: http://www.united4health.eu/.

Address for correspondence

University of Agder, PB 422, N-4604 Kristiansand, Norway.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-893

Enhancing tele-collaboration Networks by Patient Participation

Martin Staemmler[a], Heiko Münch[b], Uwe Engelmann[b], Johannes Sturm[c]

[a] *University of Applied Sciences, Stralsund, Germany*
[b] *Chili GmbH, Dossenheim, Germany*
[c] *Akademie der Unfallchirurgie GmbH, München, Germany*

Abstract

This paper decribes an approach for extending tele-collaboration to the patient and allowing the patient's participation by accessing and providing data and thereby keeping the responsibilty for maintaining a personal health record with the patient. The approach has been implemented and is in use with the nationwide tele-collaboration network TKmed.

Keywords:
patient participation; privacy; personal health record

Introduction

Conventional tele-applications like teleradiology or telepathology involve professional users in a doctor to doctor setting. Nearly all of the tele-applications lack the involvement of the patient. On the one hand these include access to the data for viewing and downloading and on the other hand the ability to upload and provide data to institutions or organisations, respectively doctors, as preparation for a patient visit or consultation for example. This poster presents an approach for extending existing tele-application systems through a patient-centric functionality without compromising current use and in particular privacy and data security. To illustrate the approach the established tele-collaboration system TKmed [1] is used.

Methods

TKmed is a nationwide tele-collaboration service currently being used in more than 120 hospitals and practices mainly in Germany and in some of the adjacent nations in cross-border scenarios. It supports tele-applications like teleradiology and telecardiology and includes a consulation workflow support.

Two patient-centric use cases have been identified by consulting patients: (i) Data object uploading for preparing a patient visit to a practice or hospital and (ii) viewing and downloading of data objects in the context of a patient visit. From an IT point of view there is an unidentified and unregistered patient requirement of access to a highly secured tele-collaboration infrastructure. Allowing a direct one step upload to registered users is prone to spam. Instead a two step upload is needed including a light weight registration by entering the patients email address and selecting the institution in the first step. Having activated the link received by email the user selects data objects and starts the upload. Data objects can be ordinary files of any type as well as individual DICOM objects from a CD or organized by a DICOMDIR.

The second use case is limited to the patient as a recipient thereby relying on a simplified consent based on the doctor-patient interaction during treatment with an intrinsic identification. To assure that only the patient is able to access data, a two-factor authentication has been established based on knowledge (login and password for the patient's email account for receiving a link) and a token generated by a TKmed application and printed at the institution (ownership) and handed over to the patient. The token is the result of a hash of demographic data and a secret provided by Tkmed at running time.

Results

Both use cases have been implemented in TKmed using a Java application being executed in the browser. The services were released to several hospitals at the end of 2014. The evaluation will be based on online questionnaires sent to the users assessing the usability of the applications and feedback regarding this support for patient participation. The first evaluation results will be available in time for the Medinfo 2015 conference.

Discussion

Exchange of information between patients and professionals depends on the eHealth literacy of the healthcare system and the institutions involved. Well-advanced, electronic record based environments support mutual access but require a sophisticated infrastructure for identity and access management based on a public key infrastructure and token. Enhancing an established tele-collaboration system is a lightweight and targeted way to achieve direct acceptance by clinical users. Patient participation has proven its usefulness for supporting patient compliance and also in taking an active role to maintain and improve the personal health status [2]. Providing such services by healthcare organisations offers a new way and has good chances of improving and fostering the relationship with the patient.

References

[1] Staemmler M, Walz M, Weisser G, Engelmann U, Weininger R, Ernstberger A, Sturm J, Establishing End-to-End Security in a Nationwide Network for Telecooperation. Amsterdam: IOS Press, 2012; pp. 512-516.

[2] Walker J, Meltsner M, Delbanco T, US experience with doctors and patients sharing clinical notes, BMJ, 2015; pp. 350-352.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-894

The Effect of Mobile App Follow-up Care on the Number of In-person Visits Following Ambulatory Surgery: A Randomized Control Trial

Kathleen Armstrong[a,b], Peter Coyte[b], John Semple[b,c]

[a] *Division of Plastic & Reconstructive Surgery, University of Toronto, Toronto, ON, CAN*
[b] *Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, CAN*
[c] *Department of Ambulatory Surgery, Women's College Hospital, Toronto, ON, CAN*

Abstract

Women's College Hospital (WCH) in Toronto offers specialized ambulatory surgical procedures. A feasibility study using a mobile appliciation (app) to supplement in-person follow-up care after surgery suggests that the mobile app adequately detects postoperative complications, eliminates the need for in-person follow-up care and is cost-effective. This is concordant with other postoperative telemedicine studies.

The purpose of this study is to determine if we can avert in-person follow-up care through the use of mobile app compared to conventional, in-person follow-up care in the first month following surgery amongst breast reconstruction patients at WCH. This will be a pragmatic, single-centre, open, controlled, 2-arm parallel-group superiority randomized trial.

Mobile app follow-up care is a novel approach to managing patients postoperatively with the potential to avert in-person follow-up and generate cost-savings for the healthcare system and patient.

Keywords:

mobile app; ambulatory surgical procedures; postoperative care; telemedicine; health services research

Introduction

Women's College Hospital (WCH) offers specialized ambulatory surgical procedures. Patients often travel great distances to undergo surgery. Most patients receiving ambulatory surgery have a low rate of postoperative events necessitating clinic visits. Regular follow-up is still considered important in the early post-operative phase. Telemedicine data suggests that mobile monitoring and follow-up care is valued by patients and can reduce costs to society [1,2].

In a feasibility study at WCH, a mobile app (QoC Health Inc., Toronto) was used to supplemement in-person postoperative follow-up care in breast reconstruction patients. Preliminary studies suggest that mobile app follow-up care is feasible, can avert in-person follow-up care and is cost-effective from a societal and health care system perspective [3].

Materials and Methods

This is a pragmatic, single-centre, open, controlled, 2-arm parallel-group superiority randomized trial comparing mobile app and in-person follow-up care over the first month following surgery. Permutated-block randomization was conducted by blocks of 4-6 using the program ralloc in Stata 13.1 (2014, StataCorp, TX).

Patient Population: All postoperative ambulatory surgery patients undergoing breast reconstruction at WCH.

Intervention: Post-operative mobile app follow-up care using the quality of recovery-9 (QoR9) and a pain visual analog scale (VAS), surgery-specific questions, surgical site photos, and frequency of submissions as defined by the surgical team.

Outcomes: The primary outcome is the total number physician visits related to the surgery over the first month postoperative. Secondary outcomes include: 1) the total number of phone calls and emails to a health care professional related to surgery; 2) complication rate; 3) societal and health care system costs; and 4) patient satisfaction over the first month postoperative.

Sample Size: If we assume that at least one in-person visit is averted in the mobile app arm, an E-test for count data generates a sample size of 70 (35 patients per group) at a power of 95 percent (p=0.05) with a 10% drop out rate.

Analysis: Count variables will be analyzed using poisson regression. Categorical variables will be tested using a chi-square test. Cost-effectiveness will be measured using net benefit regression.

Results

Results available in June 2015.

References

[1] Hwa K, Wren SM. Telehealth follow-up in lieu of postoperative clinic visit for ambulatory surgery: results of a pilot program. JAMA Surg. 2013 Sep;148(9):823–7.

[2] Gray RT, Sut MK, Badger S, Harvey CF. Post-operative telephone review is cost-effective and acceptable to patients. Ulster Med J. 2010 May;79(2):76–9.

[3] Armstrong KA, Semple JL, Coyte PC. Replacing ambulatory surgical follow-up visits with mobile app home monitoring: modeling cost-effective scenarios. J Med Internet Res; 2014 Sept;16(9):e213.

Acknowledgments

This project is funded by Canadian Institutes of Health Research e-Health Catalyst Grant.

Address for correspondence

76 Grenville St, 6th Floor, Toronto, Ontario, Canada M5S 1B2

Conflict of Interest Declaration

Dr. Semple is a shareholder QoC Health Inc.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-895

Impact of a mobile health aplication in the nursing care plan compliance of a home care service in Belo Horizonte, Minas Gerais, Brazil

Felipe A. de Britto[a], Tatiana B. Martins[b], Gustavo A. P. Landsberg [c]

[a] *Health Information Technology Management of Unimed BH, Belo Horizonte, Brazil*
[b] *Home Care Management of Unimed BH, Belo Horizonte, Brazil*
[c] *Health Information Technology Management in Unimed BH, Belo Horizonte, Brazil*

Abstract

Objective: To assess impact of a mobile health solution in the nursing care plan compliance of a home care service. Methods: A retrospective cohort study was performed with 3,036 patients. Compliance rates before and after the implementation were compared. Results: After the implementation of a mobile health aplication, compliance with the nursing care plan increased from 53% to 94%. The system reduced IT spending, increased the nursing team efficiency and prevented planned hiring. Conclusion: The use of a mobile health solution with geolocating feature by a nursing home care team increased compliance to the care plan.

Keywords:

Home Care Services; Patient Care Planning.

Introduction

Managing a home care team and evaluating compliance to the nursing care plan has always been a difficult task, due to the scope, geographical distribution and number of professionals involved. Studies show that the frequency of home visits by health professionals significantly impacts in hospital-free time of the patient [3]. The objective of this study was to assess the impact of a mobile health solution on the nursing care plan compliance by a home care service.

Methods

A retrospective cohort study was performed with the data from 3,036 patients treated between December 2013 and November 2014 by the Unimed-BH Home Care Program. All patients treated by the nurses with a Case Management care plan during this time period were considered eligible. The information on patients whose nurse visit interval was shorter than 35 days was extracted from the database of the care/administrative system, and compliance to the nursing care plan before and after the mHealth tool implementation was compared.

Results

Before the implementation of the mobile solution, care plan compliance rates were assessed quadrimestrally and were as low as 53% from January to April 2013 and 54% from May to August 2013. After the first phase of implementation, in November 2013, nursing care program compliance rates were vastly improved, increasing to 78%. After the implementation of the second phase in August 2014, nursing care plan compliance reached 94% between September and November 2014. During this period there was a small decrease of average portfolio patients compared with the previous three months (Table 1). This decrease was the result of hiring new workers in August and September. The company had forecasted nurse hirings for October and November in order to match the increasing number of patients, but as the rate of nursing care plan compliance reached satisfactory levels, the hiring of additional nurses was not deemed necessary.

Table 1– Compliance with care plan after the implementation

Period	Compliance with care plan	Average patients in the portfolio	Average daily visits
Dec/13- Feb/14	78%	65	3.28
Mar/14-May/14	79%	65	3.32
Jun/14-Aug/14	81%	64	3.49
Sep/14- Nov/14	94%	63	3.59

Discussion

The implementation of a mobile solution made clinical data in-puts easier and clinical nursing processes simpler to control. It increased compliance with the plan of care indicators, reduced IT spending, increased the team average production capacity and prevented planned hires.

Conclusion

The increasing usage of home care programs as an alternative to hospital-centered model creates complex problems that require new solutions. The use of mobile health tools may help to increase the compliance to the care plan by making patient's portfolio easier to manage and enable better planning for future visits.

References

[1] Bisione F.M., et al. Cad. Saúde Pública vol.29 supl.1 Rio de Janeiro: Avaliação da efetividade da atenção domiciliar de uma cooperativa médica de Belo Horizonte, Minas Gerais, Brasil, 2013.

Address for correspondence

Felipe Araújo de Britto – felipebritto@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-896

Augmented Reality: Real-Time Information Concerning Medication Consumed by a Patient

Gloria Diodati[a], Adrián Gómez[a], Marcela Martínez[a], Daniel Luna[a], Fernán González Bernaldo de Quiros[a]

[a]*Departamento de Informática en Salud, Hospital Italiano de Buenos Aires, Argentina*

Abstract

This paper describes a mobile prototype capable of recognizing characters from a photograph of a medication package. The prototype was built to work on the iOS platform and was developed using Objective-C and C programming languages. The prototype, capable of recognizing text out of an image, included image processing algorithms, text processing algorithms, and techniques to search and handle information from a database. This prototype is presented as an option for capturing reliable and validated information by using new technologies available to the general population.

Keywords:

Personal Health Record, Mobile Health, Optical Character Recognition, Image Processing.

Introduction

Self-medication is common, so a process for registering medications used by a patient is essential for finding discrepancies and improving the effectiveness and safety of treatments. Personal health records (PHR) facilitate doctor-patient communication and allow the patient to participate actively in self-care.

The objective of this paper is to describe the tools and algorithms used to build a mobile prototype that can recognize text from a photograph of a medication package. Integration between the PHR and this prototype is not described in this paper since the prototype is still in the conceptual definition stage. Likewise, the prototype has not been tested with real patients.

Materials and Methods

A prototype was built to work on the iOS platform and developed using Objective-C and C programming languages.

Results

During the first phase of this work the image was resized to 640 by 640 pixels. Next, the image was converted to grayscale and an engine named Tesseract was invoked to recognize characters as text. With the data obtained from this procedure, all the characters that did not belong to the alphanumeric set were removed and all words containing fewer than three characters were ignored. Then, the Levenshtein distance between the recognized words and the ones in a product database was calculated in order to select those with the smallest possible edit distance. Following this procedure, the product and its main components were identified.

Table 1 – Accuracy after first phase

Complete	Partial	Error
41.66%	6.25%	6.25%

Table 2 – Average time to process images recognized during first phase

	First processing	Call to Tesseract engine	Text processing	Text validation	Database search
Time (s)	0.0435	0.4951	0.0055	0.8637	0.1573

In the second phase the images were processed with a thresholding algorithm as a segmentation method. The threshold value used for this task was the mean value of the pixels of each image. Once preprocessing of the image stage finished, the Tesseract engine was invoked again and the text processing method was repeated. With this information we looked up the product in a database and got the pharmaceutical information.

Table 3 – Accuracy after second phase

Complete	Partial	Error
33,33%	2.08%	8.33%

Table 4 - Average time to process images recognized during second phase

	Second processing	Call to Tesseract engine	Text processing	Text validation	Database search
Time (s)	0,0499	0,3201	0,0648	0,5842	0,2505

Conclusion

Patients' communication with the doctor about medicine consumption is necessary for the doctor to make accurate diagnoses. This prototype can capture medication information using new technologies available to the population.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-897

Description of a Mobile-based Electronic Informed Consent System Development

Min-A Hwang, In Ja Kwak

Information Office, Seoul National University Hospital, Korea

Abstract

Seoul National University Hospital constructed and implemented a computer-based informed consent system in December 2011. As of 2013, 30% of the informed consents were still filled out manually on paper. Patients and medical staff continuously suggested the implementation of a system for electronic informed consent using portable devices. Therefore, a mobile-based system for electronic informed consent was developed in 2013 to prevent the issues that arise with computer-based systems and paper informed consent. The rate of filling out electronic informed consent increased from 69% to 95% following the implementation of the mobile-based electronic informed consent. This construction of a mobile-based electronic informed consent system would be a good reference point for the development of a mobile-based Electronic Medical Record and for various mobile system environments in medical institutions.

Keywords:
Computer Systems; Informed Consent; Mobile Apps

Introduction

In December 2011, Seoul National University Hospital developed a system where patients could fill out electronic informed consent to prevent the issues that arise with paper informed consent. However, paper informed consents were unavoidable when the patients had motor coordination problems, or they had to fill out the informed consent at a location where a computer was not set up.

Therefore, the percentage of those who filled out the electronic informed consent was only about 69%, on average. Further, the paper informed consents were available by default, so some of the required responses were not filled out; therefore, the main issue with using paper informed consent was still present. To avoid this, the researchers developed a mobile-based system for electronic informed consent.

Methods

The system for mobile-based electronic informed consent was constructed from June 2014 to August 2014. Departments and medical staff that usually have their patients fill out informed consent worked together on this project. The mobile devices used in the electronic informed consent development were selected based on the following criteria:

- Easy-to-read screen, Portable size and weight
- Drawing feature and Durable
- Affordable

The following features of the mobile-based electronic informed consent were selected and applied:

- Included drawing features
- Supported certified electronic signature modules
- Prevented forgery by saving captured images

Patients and medical staff in all departments, emergency rooms, and examination rooms began filling out electronic informed consent using mobile devices in September 2014.

Results

The rate of filling out electronic informed consent increased from 69% to 95% following the implementation of the mobile-based electronic informed consent. Thirty-five percent of the electronic informed consent were mobile-based. Scanning paper informed consent decreased to 50%, and the mobile electronic system prevented the respondents from missing required responses.

Interns and medical doctors were mainly the ones who filled out the informed consent, and they were interviewed regarding their feedback on the mobile-based electronic informed consent. They gave mostly positive feedback on the portability of the device, but they offered negative feedback concerning the following features:

1. Using fingers to draw and write letters is not user-friendly.
2. Takes a longer time to fill out responses on the mobile device than on paper.
3. Wi-fi can be unstable at times.

Conclusion

Implementing mobile-based electronic informed consent decreased the scanning work and lowered the required response miss rate to 0%. However, we cannot assume that these responses were more accurate, or that the questions were written more thoroughly. Quality evaluation of the electronic informed consent will be continuously out forward to secure the patients' right to know and the medical staff's responsibility to explain everything thoroughly to the patients.

References

[1] Kim Y, Kim SS, Kang S, Kim K, Kim J. Development of mobile platform integrated with existing electronic medical records. Healthc Inform Res. 2014 Jul;20(3):231-5.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-898

Exploring the Challenges and Opportunities of eHealth Tools for Patients with Sickle Cell Disease

David-Zacharie Issom[a], Alexandra Zosso[b], Frederic Ehrler[c], Rolf Wipfli[c], Christian Lovis[c], Sabine Koch[a]

[a]*Department of Learning, Informatics, Management and Ethics, Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden*
[b]*Institute of Socio-Economics, Faculty of Sciences of the Society, University of Geneva, Geneva, Switzerland*
[c]*Division of Medical Information Sciences, University Hospitals of Geneva, Geneva, Switzerland*

Abstract

Sickle cell disease (SCD) is the most prevalent inherited blood disorder in the world. The symptoms decrease the quality of life of patients and can cause premature death. Self-help solutions for chronic patients are rising and increase the quality of life of patients. We are interested to identify the usefulness of eHealth tools for patients with SCD. We did a literature review to identify the main problems faced by patients and the existing eHealth solutions. The results show a low number of studies in the field but a multi-disciplinary interest. Positive health benefits for patients are reported as well as the need for more research. Patients and caregivers lack of education about the disease, healthy behaviors are unknown and clinical best practices underused. E-health tools can offer an appropriate support for the self-management of SCD by improving the quality of life of patients, by enhancing patient health literacy and by allowing clinicians to make better decisions.

Keywords:

Sickle Cell Disease, eHealth, mHealth, telemedicine, self-care, self-management, self-monitoring, wearable devices.

Introduction

Sickle Cell Disease (SCD) is a complex hereditary form of anemia affecting millions of people throughout the world. Patients mostly suffer from painful and debilitating vaso-occlusive crisis (VOC) lasting around 7 days. Usually, they are managed at home, and patients seek hospitals only when the pain cannot be controlled anymore or when life threatening complications occur. Patients can reduce the occurence of crisis with a good self-care. For instance by following a healthy diet, keeping a moderate body temperature, avoiding tiredness, low oxygen level and by taking appropriate medicine. Consequently, eHealth tools have an important role to play, as demonstrated with the self-management of patients with diabetes [1].

Methods

We performed a literature review to identify the main problems faced by patients and the existing e-Health solutions. We used the term “sickle cell disease” with the terms “system”, “mobile”, “cellular”, “self-management”, ”mHealth”, *OR* ”telemedicine” and searched in the usual data sources.

Results

We identified 15 relevant studies and listed them in Table 1.

Table 1 – Relevant studies of eHealth solutions for SCD

N°	Reference
1, 2, 3	Woods et al. 1998, 1999, 2000
4	Breslauer et al. 2009
5	Jacob & Gerla 2012
6	Jacob et al. 2012
7	Venugopalan et al. 2012
8	Youngchan Kim & YongKeun Park 2012
9	Cheng et al. 2013
10	Jacob, Duran, et al. 2013
11	Jacob, Pavlish, et al. 2013
12	Durfee et al. 2014
13	Gonzalez-Hidalgo et al. 2014
14	Panneerselvam 2014
15	Shah et al. 2014

Discussion and Conclusion

The low number of studies indicates little interest in the field but some papers proved the importance of using eHealth tools for SCD and showed various solutions. Patients often suffer from a lack of coordinated care and from a bad understanding of their needs. This is due to the relative uncommonness of the disease. Limited financial resources, poor patient health-literacy and existing beliefs can also prevent patients from making the most valuable health choices. Therefore, using eHealth to provide trustworthy information is recommended. The reviewed papers show that eHealth tools can be useful for self-monitoring purposes. Few work has been actually done, but the potential of improving health outcomes is great. Because the disease is often not diagnosed, mobile technologies can facilitated the screening and the diagnosis phases. A change in patients’ behavior can be influenced with mobile applications helping them to adopt and follow healthy practices, such as with daily advices or notifications. Further studies need to assess which technologies are able to tackle the main unmet patient needs.

Reference

[1] Årsand E, Tatara N, and Hartvigsen G, “Wireless and Mobile Technologies Improving Diabetes Self-Management,” pp. 136–156, 2011.

Address for correspondence

David-Zacharie Issom,
CH-1208 Geneva
david.issom@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-899

Development and Usability Evaluation of the Mobile Delirium Assessment App Based on Confusion Assessment Method for Intensive Care Unit (CAM-ICU)

Meihua Ji[a], Ying Wu[a], Polun Chang[b], Xin Yang[c], Fangyu Yang[a], Shuang Xu[a]

[a] *School of Nursing, Capital Medical University, Beijing, China*
[b] *Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan*
[c] *Affiliated Xuanwu Hospital, Capital Medical University, Beijing, China*

Abstract

Delirium is a common complication among patients in ICU settings. The accuracy of using the assessment tool CAM-ICU to detect delirium is relatively low during routine practice among bedside nurses. The aim of this study is to develop a mobile application (app) to detect delirium in early stage and to test its usability among ICU nurses. The app was developed with Java and installed on a mobile device. A questionnaire was created based on the Technology Acceptance Model (TAM) measuring their response to the four domains of TAM: perceived usefulness (PU), perceived ease of use (PEOU), attitudes towards usage (ATU) and behavioral intention to use (BIU). One hundred and two ICU nurses completed the survey. The result indicated that the app we developed has easy to use interfaces and is easier to use compared to the regular CAM-ICU.

Keywords:
Mobile Application; CAM-ICU; Delirium.

Introduction

Delirium is one of the most common complications of ICU patients and can adversely affect patients' short and long-term outcomes. It should be monitored regularly using CAM-ICU as reccomended by American College of Critical Care Medicine. Although the acuracy of CAM-ICU in detecting delirium is reported repeatedly high (> 90%) in research papers, its sensitivity and specificity are only around 50% when it was used in routine nursing practice due to lack of adequate training, inaccurate and incomplete interpretation of assessment data or using variable patient baselines when making the diagnosis [1]. The purpose of this study is to develop a standardized and easy to use mobile CAM-ICU delirium detecting app for bedside nurses and test its usability.

Materials and Methods

This study includes two phases: development of the mobile app and evaluation of its usability. The app was designed with JAVA and installed on a mobile device. A modified usability evaluation questionnaire of TAM, which includes four domains and 46 items and rated with 5-Likert scale, was used to test the usability of the developed app. There are 4 items comparing the level of easiness between the mobile app and regular CAM-ICU. Convenience sampling was used, and nurses from ICU settings of three hospitals completed the questionnaire after using both mobile and regular CAM-ICU tools.

Results

The screenshots of the app are shown in Figure 1. The app was programmed to automatically retrieve baseline data for comparision and print out the result on whether or not the patient has delirium.

A total of 102 ICU nurses participated to test the usability of the mobile CAM-ICU app. The means of the four domains regarding to the app are: PU: 4.09-4.22, PEOU: 4.07-4.33, ATU: 3.90-4.16, BIU: 3.97-4.24. The Cronbach's α of each demain is 0.936, 0.949, 0.902, 0.915. The means of the four items that compare the mobile app and regular CAM-ICU are 4.14, 4.17, 4.23, 4.23, indicated that nurses have perceived the app as easier to use compared to the regular CMA-ICU.

Figure 1. Application Screenshots

Conclusion

The mobile CAM-ICU app we developed has easy to use interfaces. The evaluation indicates that this is a useful tool in delirium assessment and easier to use compared with regular CAM-ICU, especially in guiding nurses through assesment accurately and observing patient's condition comprehensively.

References

[1] Reade M C, Eastwood G M, Peck L, Bellomo R & Baldwin I. Routine use of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) by bedside nurses may under diagnose delirium. Critical Care and Resuscitation, 2011, 13: 217-224.

Address for correspondence

Ying Wu, helenywu@vip.163.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-900

Viability of a Bioelectrical Signal Acquisition System Energized by Cellphone with NFC

Marcel Seiji Kay, Fábio Iaione

Faculty of Computing, Federal University of Mato Grosso do Sul, Campo Grande

Abstract

Currently, smartphones are used in various systems in the medical field due to the presence of various features, notably Near Field Communication (NFC). NFC utilizes communication technology and an energy supply based on electromagnetic induction. One of the most common medical tests is the electrocardiogram (ECG), through which various heart diseases can be diagnosed. The objective of this study is to evaluate the feasibility of providing power to a bioelectrical signal acquisition module using a mobile phone with NFC. After testing it was indicated that it is possible to construct a passive module to acquire ECG signals using NFC mobile phone.

Keywords:

Electrocardiogram; Near Field Communication; M-health.

Introduction

The use of passive devices in the medical field is a relatively common practice, especially when they need to be implanted [1], where having no battery is a very interesting feature. In passive systems, data communication and transfer of energy occur through inductive coupling between planar coils [2]. The energy transfer by this method is approximately 80% at a distance of 15 mm [3]. The depth of penetration of electromagnetic waves on living tissue decreases significantly with increasing frequency, so the frequency of the Radio Frequency Identification (RFID) systems used in living tissue is between 125 KHz and 40 MHz [2,4]. Smartphones have great processing power, storage, visualization and communication, including Near Field Communication (NFC), and are already used in the medical field. One of the most commonly performed medical tests is the electrocardiogram (ECG), which can diagnose various heart diseases, thus enabling a pretreatment to prevent possible sequelae. Accordingly, the objective of this study is to assess the feasibility of providing power to a bioelectrical signals acquisition module using a mobile phone with NFC.

Methodology and Results

In this study a reimplementation of the bioelectric signals acquisition module described in [5] was used. This module performs conditioning, digitization and transmission of the EEG signal captured. It can also capture the ECG signal reducing the gain of the amplifiers. The module is powered by a battery whose voltage is regulated to 5V, with a current of 17 mA (P = 85mW). The transmission of the digitized signal is performed by serial infrared at 9600 bps. Table 1 shows the electric current of several components.

A test was conducted using a mobile phone, Samsung model GT-I9250, and a rectangular coil of four turns (26 AWG), measuring 4 x 8 cm. In this test, loading of 1500 ohms was measured as a sinusoidal signal (13.56 MHz) of 10 Vpp (peak-to-peak voltage), indicating transferred power of 8.33 mW.

Table 1 – Components of the module and their currents

Component	Current
Instrumentation Amplifier (AD627) and Operational Amplifier (TS912)	785 uA
Microcontroller (AT89C2051 @ 11.092 MHz) and ADC (MAX187)	12 uA
Voltage regulators (2 x LP2950-50)	400 uA
Infrared serial communication (LED)	4 uA

Conclusion

Exchanging the microcontroller by C8051F320 (internal ADC) which consumes 0.6 mA (at 1 MHz), and removing the infrared communication, the power module can be reduced to approximately 8.9 mW (5V x 1.785 mA), which is compatible with the measured transferred power. Furthermore, it is also possible to optimize the number of coil turns, increasing the power transferred. Therefore, it is possible to construct a passive module for acquisition of bioelectrical signals without energy source, using an NFC mobile phone to display the signal and power supply.

References

[1] Hara MS, Abatti PJ. Two components implantable device for passive thermotelemetry. In:Advance in Medical Diagnos- tic Techniques in Procedures. New Delhi, 2004, pp. 106-109.

[2] Dobkin D. The RF in RFID passive UHF RFID in practice. Burlngton: Elsevier, 2008.

[3] Si P, et al. A frequency control method for regulating wireless power to implantable devices. IEEE Transactions on Biomedical Circuits and Systems, 2008: 2: 22-29.

[4] Oberle M. Low power systems-on-chip for biomedical applications. Konstanz: Hartung-Gorre Verlag, 2003.

[5] Fábio I, Jefferson M. Methodology for hypoglycaemia detection based on the processing, analysis and classification of the electroencephalogram. Medical & Biological Engineering & Computing, 2005: 43 (4): 501-507.

Address for correspondence:

marcel.kay@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-901

Smart Glasses – A New Tool in Medicine

Gunnar O Klein[a], Karandeep Singh[b], Johan von Heideken[c,d]

[a] *Dept. of Informatics, Örebro University, Örebro, Sweden*
[b] *Dept. of Medicine, Harvard Medical School, Boston USA*
[c] *Dept. of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden*
[d] *Department of Physical Therapy, Movement and Rehabilitation Sciences, Northeastern University, Boston USA*

Abstract

Smart glasses, defined as a computerized communicator with a transparent screen and a video camera, wearable as a pair of glasses, have started to be tested for a variety of health related applications. This poster reviews some of the early experiences and gives a series of proposals for possible uses in medicine with a particular emphasis on medical education.

Keywords:

Clinical informatics; Educational models; Google Glass.

Introduction

Smart glasses have the following set of features:

- A hands free communicator, that can communicate via Bluetooth with, for example, a smart phone, or directly with the Internet when Wi-Fi is available.
- A video camera and voice recorder/transmitter.
- A display for viewing text and images.
- A voice input interface.

The most published product today is Google Glass which has been available for pilot beta tests since February 2013. Like tablets and smart phones, smart glasses are a platform for various applications [1]. They were not specifically targeted at health care but there are already numerous pilots supporting various medical applications. On January 15, 2015 Google announced [2] that the production of the Glass prototype was stopped but Google remains committed to the development of the product.

Methods

This poster is based on a literature review of 17 full papers.

Results

Since their release, smart glasses have been tested in health care for a variety of applications using limited numbers of patients and reported in various news media and conferences. Our literature review found seventeen scientific papers which provide a clinical evaluation of these new devices. In addition to monoapplication papers, Muensterer *et al* had a pediatric surgeon wear a glass device throughout the day for four weeks and recorded possible situations where the technology might assist. These papers provide mostly promising results but also caution where the technology was deemed not suited for the tested application, for example for ECG interpretation and autopsy documentation.

Table 1 – Different application areas described

Application area	Findings
Remote instruction of users wearing Glass	Beneficial use in cardiologist education, Orthopedic surgery training, Cancer surgery, Central venous access, Ultrasound interpretation, and in Diabetic limb assessment
Documenting procedures	Autopsy documentation, Airway intubation
Patient empowerment	Allergy patients getting access to information, Macula patients getting augmented vision
Reading signal data	ECG assessment, Immunochromatography, Vital signs during radiological intervention
Providing instructional films and simulation	Disaster medicine, Anatomy and palpation

Discussion

There are many potentially important applications of this new technology in medical education both for remote supervision and for the production of educational films. The usefulness in clinical practice, for example for seeking a second opinion or providing assistance to the workflow or reading of adapted EHRs remains to be studied.

References

[1] Google Glass: Tech specs. Available at: https://support.google.com/glass/answer/3064128?hl=en. Accessed 6th, August, 2014.
[2] Google Glass sales halted but firm says kit is not dead. Available at: http://www.bbc.co.uk/news/technology-30831128. BBC News. January 15, 2015.

Address for correspondence

Professor Gunnar O Klein, Centre for Empirical Research on Information Systems, Örebro University Business School, Fakultetsgatan 1, 701 82 Örebro, Sweden. E-mail: gunnar.klein@oru.se

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-902

Access Control for Mobile Assessment Systems Using ID

Masaharu Nakayama[a], Tadashi Ishii[b], Kazuma Morino[c]

[a] *Department of Disaster Medical Informatics, International Research Institute of Disaster Science, Tohoku University, Miyagi, JAPAN*
[b] *Department of Education and Support for Community Medicine, Tohoku University Hospital, Miyagi, JAPAN*
[c] *Emergency Medicine, Yamagata Prefectural Central Hospital, Yamagata, JAPAN*

Abstract

The assessment of shelters during disaster is critical to ensure the health of evacuees and prevent pandemic. In the Ishinomaki area, one of the areas most damaged by the Great East Japan Earthquake, the highly organized assessment helped to successfully manage a total of 328 shelters with a total of 46,480 evacuees. The input and analysis of vast amounts of data was tedious work for staff members. However, a web-based assessment system that utilized mobile devices was thought to decrease workload and standardize the evaluation form. The necessary access of information should be controlled in order to maintain individuals' privacy. We successfully developed an access control system using IDs. By utilizing a unique numerical ID, users can access the input form or assessment table. This avoids unnecessary queries to the server, resulting in a quick response and easy availability, even with poor internet connection.

Keywords:

Disaster; Shelter; Mobile; Assessment.

Introduction

The magnitude-9.0 Great East Japan Earthquake struck the northeastern coast of Japan on the 11th March 2011, resulting in 15,889 dead, 2,597 missing, and 6,152 injured persons, as of August 2013. During the disaster, many people (around 400,000) were evacuated to temporary shelters, such as public halls, gymnastic halls, and schools in Northeast Japan. One of the most damaged areas was Ishinomaki, which had a total of 328 shelters with a total number of 46,480 evacuees. Cooperating with various pertinent institutions—such as the Miyagi Prefecture government, the Ishinomaki municipal government, the local medical association, the Tohoku University Hospital, and the Japan self-defense forces—the Ishinomaki Zone Joint Relief Team consisted of physicians, nurses, public health nurses, pharmacists, and volunteers that were based in the Ishinomaki Medical Zone. This team managed a total number of 53,696 persons in evacuation shelters for almost six months [1]. During the management of disaster relief, the team continued a survey of medical and sanitary conditions, as well as environmental conditions, including power, gas, and water supplies. Their efforts resulted in successful and well-organized management, avoiding the development of pandemic. However, the process of evaluation, calculation, and analysis was associated with tedious work and high workloads for staff members. To decrease this burden, an assessment system that utilizes mobile devices was developed. This study details one of the requirements for the system and the means of meeting this requirement.

Methods

The software is a browser-based app that is available on the iPad. The database on the server is based on MongoDB. Since Internet availability is often poor under disaster conditions, the input data are able to be stored off-line and are sent to the server after re-connection to the Internet.

Results

The software has two main functions: data input and data management. A medical crew is required to input data on about 19 assessment factors, describing the current condition of shelters in the area, into the mobile assessment system. The headquarters of the area analyzes the collected data from the dispatched relief teams using assessment tables, contained within the mobile assessment system. In relation to both privacy and efficiency, controlled access to the system is required. The relief teams are limited to inputting data about the shelters that they are in charge of. The headquarters of the areas are then able to browse information on the shelters that they are charged with managing. Furthermore, senior command teams are able to access the information from a wide range of areas. These access control practices utilize ID numbers. For example, if 99 relief teams have IDs 101 to 199, then a relief team with ID 101 can access the information about shelters linked to ID 101, but cannot access information linked with other IDs. Furthermore, the headquarters with ID 100 that is in charge of crews with IDs 101 to 199 can access all the available information about the shelters linked to those IDs.

Conclusion

Our mobile assessment system is available for both headquarter and dispatched relief teams and utilizes access control with ID numbers. This system is useful for avoiding unnecessary queries to the server when internet connections are poor.

Acknowledgments

We are greatly thankful for the technical support of Jun Yamadera and Koichiro Amito (Eyes, JAPAN Co. Ltd.).

References

[1] Ishii T. Medical response to the Great East Japan Earthquake in Ishinomaki City. Western Pacific Surveillance and Response Journal 2011;2(4):10–16.

Address for correspondence

1-1 Seiryomachi, Aobaku, Sendai, Miyagi, JAPAN 980-8574

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-903

Using mobile devices to improve the safety of medication administration processes

Navas H[a], Graffi Moltrasio L[a], Ares F[b], Strumia G[c], Dourado E[a], Alvarez M[a]

[a] *Sanatorio Finochietto, Buenos Aires, Argentina*
[b] *Develaris, Buenos Aires, Argentina*
[c] *TIPS Salud, Córdoba, Argentina*

Abstract

Within preventable medical errors, those related to medications are frequent in every stage of the prescribing cycle. Nursing is responsible for maintaining each patients safety and care quality. Moreover, nurses are the last people who can detect an error in medication before its administration. Medication administration is one of the riskiest tasks in nursing. The use of information and communication technologies is related to a decrease in these errors. Including mobile devices related to 2D code reading of patients and medication will decrease the possibility of error when preparing and administering medication by nurses. A cross-platform software (iOS and Android) was developed to ensure the five Rights of the medication administration process (patient, medication, dose, route and schedule). Deployment in November showed 39% use.

Keywords:

Electronic Health Record; Mobile Applications; Medication Errors; Patient Safety.

Introduction

The prescribing cycle, which consists of five steps since the treatment is indicated until it becomes effective, contains errors at each step. The most common errors in medication preparation and administration are: dosage, frequency, route of administration, patient, starting time or suspension, medication, no medication administration, wrong infusion preparation, mishandling of infusion pumps, or inadequate record.

The nurse is the last person who can check if the medication was correctly ordered and dispensed before administration. Medication administration is probably the highest risk task a nurse can perform. Applying the Rights to medication administration can enhance patient safety and prevent harm. The use of barcodes during medication administration reduces the possibility of error. Our objective is to describe the mobile system developed and its inclusion in a safe medication process.

Methods

Sanatorio Finochietto is a private healthcare institution belonging to the *ASE (Acción Social Empresaria)* group, founded in November 2013, with over 150 adult inpatient beds and over 20 neonatal intensive care incubators, it has 6 intelligent general operating, 2 outpatient operating, 2 delivery and 2 antepartum rooms. It is the first bio-eco-smart medical center with a structure designed for a rational and responsible use of natural resources. It possesses a commercial Health Information System (HIS).

Nurses prepare the schedule for each indication in the "nursing Kardex". Each nurse reads the medical indication of each patient, opens the medication drawer of said patient, separates the indicated medications and places them on a tray. They then go to the patient's room with the prepared tray and administer the medication. Once all the medication has been administered to the patient, nurses return to the office to record the administration of the medication.

Results

A native application for mobile devices (iOS and Android) has been developed in .NET, deployed within the aforementioned system supporting the five Rights:

- Right time: The system issues a visible and audible alert (push), and enables the preparation of the medication from one hour before the scheduled timetable by nursing, until one hour later.
- Right dose and Right medication: The system displays the scheduled medication list pertaining to a certain timetable for the selected patient. As the user selects each medication from the list, the system enables the device camera to verify the code on the medication packaging.
- Right patient: Upon completion of a patient's medication preparation, the system enables the device camera to allow the nurse to read the code on the patient's wristband before administration.
- Right route: The system displays the prepared medication list grouped by route. Users checking such list administer the medication and mark the medication which has been correctly administered directly from the device display.

One sample of administration activity was taken on a weekday and on a weekend day in November 2014. Resulting in weekday numbers: 743 without device, 483 (39%) with device, weekend numbers: 982 without device and 188 (16%) with device.

Conclusion

The degree of user acceptance of the new technologies is directly related to the user experience and time-saving of tasks completed with the technology. User training and support are crucial to enhance acceptance, as well as support given to those users who regularly use the device. We must include a nurse who has the ability to lead development and chamption implementation on the team. The institution is still in the implementation stage this technology, and until this stage is completed, is it impossible to measure whether the use of this technology actually improves patient safety.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-904

Study of Development for RFID System to Hospital Environment

Seung Kwon Hong[a], Myung-Whun Sung[b]

[a]*Division for BioMedical Informatics, Department of Family Medicine, School of Medicine, The Catholic University of Korea, Incheon St. Mary's Hospital*
[b]*Department of Otorhinolaryngology, Head and Neck Surgery, Seoul National University Hospital*

Abstract

RFID/USN develops information systems for anytime, anywhere to anybody access Electronic Medical Records (EMR). The goal of the present study is to develop a RFID/USN-based information system for the hospital environment. First, unable to recognize, second, able to recognize as a pursuit of place and suppose the time of medical examination. A retrospective analysis of 235 RFID monitoring results, from four ENT ambulatory clinics of Seoul National University Hospital were extracted by a reader program and monitoring of RFID tag (2006.11.16~2006.12.16). RFID detection for sensing reader of this study has been put into representing "place" and "spending time" of patients for medical history taking and examination. Through the RFID of detection for specific place and spending time of medical examination, RFID/USN develops information system progressing in the EMR of hospital system.

Keywords:

RFID, Electronic Medical Record, Hospital Information System.

Methods

This study researched 162 new registered patients at otolaryngology department at Seoul National University Hospital. The system comprised of RFID technology, RFID tags, reader, tag and a Reader to receive information exchanged between the servers or network bearer to middleware formation. The RFID antenna used for this study had a frequency of 902 ~ 928MHz and 5.73 dBi, 4 watts EIRP with a high-frequency power. The treatment room, hearing laboratory room, and hospitalization Reservations room where the RFID Reader was installed was from the ground to the reader center and average was 114 cm, with the entrance door widths being a maximum of 130cm and a minimum of 87cm. A reader installed in the otolaryngology entrance hall increased the recognition rate and was installed with two pairs from the ground to the Reader center height of 135cm and 68.5cm, and width of the corridor was 230cm. A desktop type of reader was installed in the nursing room and the tags were recognized by direct contact. In this study, the patient possessed RFID tag is aware of the recognition moment when the installed antennas location and the vision of a box in the form of implementation are listed in the time order, which allows researchers to use estimate pathway of outpatient. In addition, a duplicate tag was recognized by the Reader in order to solve the problem of the the patient's gait speed within 5 seconds and to account for new understanding of the signal recognition results were excluded from the program.

Results

Table 1– Number of RFID detection according to case

	Unable-to-recognize case	Duplicate-check + able-to-estimate location	Able-to-estimate location	Able-to-estimate treatment time
Initial visit	39	123/162 (0.759)	90/162 (0.555)	17/162 (0.105)
Follow-up visit	0	73/73 (1.0)	59/73 (0.808)	14/73 (0.191)
Sum	39 (0.615)	196 (0.834)	149 (0.634)	31 (0.131)

Discussion

Hospital workers are facing a new problem called 'to adapt to change', but in the past when discussing the success of hospital information hardware was the main concern then the software but recently has focused attention on human element This study explored the potential of RFID technology applied to patients in the hospital for the first time, but in the future it may improve the recognition rate with a solution methodology for medical use even in the form of quantitative methods, as well as qualitative analysis research methodology for use in future studies.

Address for correspondence

Myung-Whun Sung : SCHOOL OF MEDICINE, THE CATHOLIC UNIV. OF KOREA, INCHEON ST.MARY`S HOSPITAL *[e-mail: skdoc@snu.ac.kr]*

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-905

Developing an Online Decision Aid for Osteoarthritis

Glenn Salkeld[a], Sally Wortley[a], David Hunter[b], Hema Umapathy[b], Jack Dowie[c]

[a] *School of Public Health, University of Sydney, Australia*
[b] *Institute of Bone and Joint Research and Kolling Institute, University of Sydney, Australia*
[c] *London School of Hygiene and Tropical Medicine*

Abstract

A decision aid for osteoarthritis was developed using the best available evidence on effect size, potential harms and self-rated performance for other attributes. The aid was developed using a multi-criteria decision analytic tool capable of combing evidence and an individual's preferences for the attributes related to treatment.

Keywords:
Decision aid; osteoarthritis; multi-criteria decision analysis.

Introduction

Decision making for osteoarthritis is complex because of marked variations between patients in weight, tolerance for physical activity, and risk for anti-inflammatory adverse events including both gastrointestinal and cardiovascular toxicity. There are also variations between patients in frequent concomitant comorbidities including depression, hypertension and/or diabetes. All of these variations can influence informed osteoarthritis management decisions. This led us to develop an online decision aid as a means of combining the best available evidence on the benefits and harms of osteoarthritis management with patient preferences. The aim of the aid is to provide an opinion on which treatment option may be best for them.

Methods

The decision aid uses a generic web-based decision-support template grounded in multi-criteria decision analysis (MCDA). The app, known as Annalisa© (AL), uses a simple expected value algorithm to calculate a score for each option. The score takes into account the individual's preferences for different criteria (stored as importance weights) and the evidence of the performance of each option on each criterion.

The aid allows for the dual personalisation of the decision in terms of both the clinical characteristics of the patient and their preferences in relation to the benefits and harms associated with the alternative treatment options. It incorporates evidence on both the benefits and the potential harms of a range of osteoarthritis management options (the 'attributes') from published evidence-based guidelines, tailoring these as closely to the specific patient as possible by information elicited about the patient. By combining this evidence with the individual's importance weights for the various outcomes, which is elicited in a graphical way at the point of decision, the best course of action for each patient will be identified on the basis of quantified scores for each option. This poster will present a summary of the systematic review of the literature on both qualitative and quantitative studies reporting on treatment preferences of patients with osteoarthritis. It will also present graphically the design and content of the osteoarthritis decision aid as a work in progress.

Results

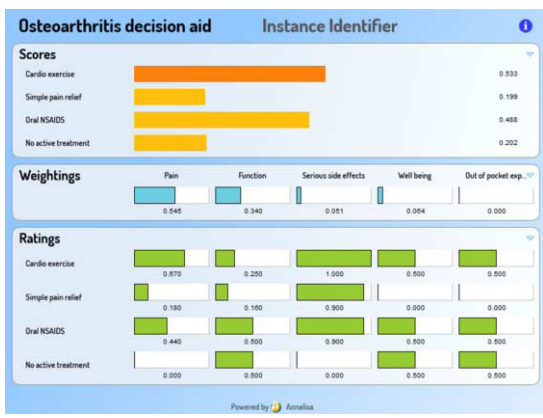


Figure 1 – Osteoarthritis decision aid

The aid will be launched in June 2015 and preliminary results will be available by July 2015. The results from the work in progress will be presented at the conference.

Conclusion

The proposed osteoarthritis decision aid uses the best available evidence to populate the aid based on the options and attributes that the user identifies as being important to them. It is feasible to combine evidence and individual's preferences in a way that provides a quantified score for determining which option is best.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-906

An electronic device to record consensual reflex in human pupil

Pinheiro H. M.[a], Costa R. M.[a], Camilo E. N. R.[b], Gang Hua[c]

[a] *Instituto de Informática, Universidade Federal de Goiás, Goiás, Brasil*
[b] *Ophthalmologist, Goiânia, Goiás, Brasil*
[c] *Department of Computer Science, Stevens Institute of Technology, Hoboken, New Jersey, USA*

Abstract

Examination of the pupil offers an objective evaluation of visual function as well as the vegetative pathways to the eye. This work proposes the development of an effective method and a portable device to test the consensual pupillary reflex. The first results demonstrate the success of a new device construction and methodology to record the consensual reflex with different stimulus, in a situation of complete blockage of light.

Keywords:
Consensual Reflex; Pupillometry; Computer-aided diagnosis.

Introduction

The pupillary inspection is a valuable part of the ophthalmological, neurological and general medical examinations routine [1]. Pupillary examination involves recording the size, symmetry, and light reactivity of both pupils [2].

Materials and Methods

To apply the test methodology, a pupilometer was built based on consensual human optics reflection. The Pupilometer has a lighting system with visible light that gradually goes from 0 (zero) to 38 lux, positioned at a 3 centimeters distance from one of the eyes and completely seals illumination as shown in Figure 1 (a). The four metrics used to test methodology were amplitude, latency, time to maximum contraction (TMC), time to maximum dilation (TMD). Twenty-nine volunteers' participated and 61 records were made, with 1 to 5 recordings for each volunteer. Eight volunteers were females (27.58%) and 21 males (72.41%). Through the software it was possible to set visual stimulus duration, intensity and choose which eye should be recorded.

Results

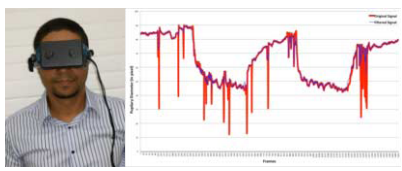
Table 1 shows the mean and standard deviation of the tests performed with 61 videos.

Table 1 – Configuration of visual stimulus

Parameter	1st Period		2nd Period	
	Mean	SD	Mean	SD
Amplitude	0.58%	0.09%	0.59%	0.08%
Latency	0.33 seg	0.57 seg	0.24 seg	0.10 seg
TMC	1.67 seg	0.37 seg	1.57 seg	0.41 seg
TMD	1.84 seg	0.29 seg	1.85 seg	0.27 seg

Discussion

In some cases, the volunteer movements caused failures on pupil segmentation and, therefore, caused noise in the signal. To correct them, the software used the neighborhood average algorithm, resulting in a sign like shown in Figure 1 (b).



(a) (b)

Figure 1 – (a) Volunteer using the pupilometer, (b) Example of the algorithm application by software

Conclusion

The Pupilometer and the software proved to be an effective, non-invasive and portable pupillary behavior after light stimulus identification method, being a useful tool to integrate in the clinical, emergency, ophthalmology and neurology related medical routine. It is necessary to conduct more clinical investigations to determine pathological patterns in patients with diseases. This work can also open a way to new studies involving computer-aided diagnosis (CAD). Changes in the software could make possible studies to identify signs of a probable disease.

References

[1] Wermund TK, Wilhelm H. Pupillary disorders - diagnosis, diseases, consequences. Klin Monbl Augenheilkd. 2010, nov;227(11):845-51.

[2] Martínez-Ricarte F, Castro A, Poca MA, Sahuquillo J, Expósito L, Arribas M, Aparicio J. Infrared pupillometry. Basic principles and their application in the non-invasive monitoring of neurocritical patients. Neurologia. 2013 Jan-Feb;28(1):41-51.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-907

A Global Analysis of Approaches to Sharing Clinical Data with Patients

Jennifer E. Prey[a], Fernanda Polubriaginof[a], Gilad J. Kuperman[a,b], Victoria Tiase[b], Sarah A. Collins[c], David K. Vawdrey[a,b]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*
[b] *NewYork-Presbyterian Hospital, New York, NY, USA*
[c] *Partners Healthcare and Harvard Medical School, Boston, MA, USA*

Abstract

Engaging patients in their care has become a topic of increasing importance, and enabling patients to have access to their clinical data is a key aspect of such engagement. To investigate on an international scale the current state of approaches for providing patients with access to their own clinical information, individuals from 16 countries, across six continents, participated in cross-sectional semi-structured interviews. Interview questions focused on social and cultural influences that affected patient engagement activities, government support for current and planned initiatives, data ownership models, and technical issues. Substantive initiatives for providing information to patients in the majority of countries interviewed are present; however, these initiatives were diverse in nature and stage of implementation. Efforts to improve patient access to data are active on a global-scale. There are many open questions about best practices and much can be learned by adopting an international perspective to guide future implementation efforts.

Keywords:

Personal Health Records; Consumer Health Information; Health Information Technology; Patients; International Perspectives.

Introduction

Around the world, people are being encouraged to participate more actively in their healthcare. While individuals in most countries have been able to obtain copies of their medical records for many years, few regularly take advantage of this opportunity. With the proliferation of electronic health records (EHRs), there is much greater potential for patients to access their information. Providing patients with their clinical data is associated with increased satisfaction, improved patient knowledge, control and self-care, and may result in better health outcomes[1].

Methods

This study was a cross-sectional semi-structured interview survey conducted via telephone or videoconference during November and December 2014. The list of interview participants was created using a convenience method of sampling in which research team members identified individuals around the world who had knowledge of patient data-sharing initiatives in their respective countries. Data from the interviews as well as information from literature review were aggregated and analyzed for themes. Countries were grouped based on similarity of characteristics and compared across themes.

Results

Interviews were conducted with individuals representing 16 countries, across six continents. Respondents reported substantive initiatives for providing information to patients in the majority of countries interviewed. There is considerable variability in the level of maturity, the degree of government involvement, the technical infrastructure, and the plans for future development across the world. As informaticians, we are still in the early stages of deploying patient engagement technologies and have yet to identify optimal strategies in this arena.

Conclusion

The sharing of clinical health information with patients is in varying stages of development around the world. The trend is toward increasing engagement of patients and providing more information, especially electronically. Based on this study it is clear that no one country has 'solved' how to provide patients with access to their clinical information. Countries should leverage the experiences of others as we all move forward in the deployment of these systems. When patients are empowered with access to their own clinical information, they will be able to more actively understand and participate in their care, and with this additional knowledge and motivation, hopefully have improved satisfaction and overall health.

Acknowledgments

This work was supported by grants from the National Library of Medicine (T15LM007079 - Hripcsak) and the Agency for Healthcare Research and Quality (R01HS21816 - Vawdrey).

References

[1] Giardina TD, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc. 2014 Jul 1;21(4):737–41.

Address for correspondence

Jennifer Prey, DBMI, Columbia University, 622 W. 168th St. PH-20, WS-14, New York, NY, 10032, USA; jep2175@cumc.columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-908

Is Access to eHealth Records Important for Patients? – Opinions of Healthcare Personnel

Isabella Scandurra[a], Ture Ålander[b]

[a] *School of Business, Örebro University, Sweden*
[b] *Department of Public Health and Caring Sciences, Uppsala University, Sweden*

Abstract

Sweden has had significant progress with the the introduction of electronic health records. A pilot county deployed in an eHealth service in 2012, giving access to health records for all of its patients. This eHealth service is, however, a controversial issue. Two surveys were conducted to discover whether healthcare professionals' opinions differ between professionals, and between staff who have had experience with patients using eHealth records and those, to date, who have had none. Experienced nurses found this eHealth service more important for the patients compared to unexperienced nurses outside the pilot county, as well as both semi-experienced physicians.

Keywords:

Electronic Health Records; Healthcare Personnel; Online Systems; Access to Information; Web Questionnaires; eHealth.

Introduction

In Sweden, patients have the right to obtain part of their health record. The national eHealth strategy states that all health records should be accessible via Internet by 2017. Within the scope of an EU-project, Uppsala County Council in November 2012, deployed such a public eHealth service to its 350 000 patients as the first national trial of patient-facing records. Access to "eHealth records" is, however, a controversial issue, as many professionals are concerned about patients reading their record without support from clinicians. Do experiences from the pilot county differ from regions where eHealth records are not yet implemented? This study aims to discover whether healthcare professionals think this eHealth service is important for the patients.

Materials and Methods

This study collected data from two 5-graded Likert scale web surveys of Swedish healthcare staff, focused on attitudes and opinions of physicians (S1) and nurses (S2). S1 was sent out in June 2013 to 1600 physicians in the pilot county with a response rate of 25% (399 respondents). S2 was sent out in March 2014 to 8460 registered nurses and midwifes in Sweden with a response rate of 35,4% (2867 respondents). The questionnaires consisted of background questions and 5 sets of items with free text fields in each set. To deliver the questionnaires, web survey tools were used.

Standard data reports were created for each survey, with charts showing the most prominent differences in each statement. Currently each statement is being analysed by statistics and healthcare informatics researchers and students. As the questionnaires were jointly developed, this first analysis is expected to discover differences between care professionals that have had experience with patients using eHealth records and those working in regions where the service was not yet implemented.

Results

The statement analysed in this study was: "To which extent patients' access to eHealth records is important for the patients?" Responses were stated on a 5-graded Likert scale from strongly disagree [1] to strongly agree [5] and are presented as a percentage, based on 241 nurses in the pilot county, all 2871 other nurses in Sweden, and 387 Physicians in the pilot county. In general, nurses' opinions of patients reading their health record online were neither entirely positive nor negative (median=3), while physicians were generally more negative (median=2).

Table 1 – "To which extent is access to eHealth records important for the patients?"

Professions	1	2	3	4	5
Nurses, pilot county	9	16	37	17	18
Nurses, All other	13	25	35	15	10
Physicians, pilot c.	28	36	27	8	2

Nurses experienced with patients having eHealth records, were more positively (35% agree and strongly agree) inclined towards this service for patients, compared to nurses in the rest of the country (25%) as well as physicians (64% strongly disagree or disagree).

Time between the questionnaires (9 months) introduces uncertainty of comparability between professions' opinions. The differences between experienced and unexperienced nurses were important to observe. The nurses in the pilot county may have experienced that access engenders a positive effect when the patient meets with healthcare providers.

Conclusion

As experienced nurses found this access more important for the patients compared to unexperienced nurses and semi-experienced physicians it is important for development to disseminate these findings to unexperiencenced peers.

Address for correspondence

Isabella Scandurra, Örebro University, School of Business, eHealth, SE-701 82 Örebro, Sweden; e-mail: isabella.scandurra@oru.se

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-909

Rolling Medical Practice: Ambulant Mobile Medical Care for Rural Areas

Jonas Schwartze[a], Klaus-Hendrik Wolf[a], Sebastian Schulz[a], Maike Rochon[b], Markus Wagner[b], Uwe Bannenberg[c], Markus Drews[a], Thomas Fischer[d], Torben Hellwig[c], Stefan Hofmann[e], Petra Höft-Budde[f], Ralf Jäger[c], Stefan Lorenz[g], Ruth Naumann[h], Maik Plischke[b], Jörg Reytarowski[f], Constanze Richter[e], Christiana Steinbrügge[i], Anja Ziegenspeck[i], Julius von Ingelheim[c] and Reinhold Haux[a]

[a] Peter L. Reichertz Institute for Medical Informatics,University of Braunschweig, Germany
[b] Braunschweig Centre of Informatics and Technology GmbH (BITZ), Braunschweig, Germany
[c] Wolfsburg AG, Wolfsburg, Germany
[d] Sozialversicherung für Landwirtschaft, Forsten und Gartenbau, Kassel
[e] Association of Statutory Health Insurance Physicians Lower Saxony, Braunschweig, Germany
[f] AOK - Die Gesundheitskasse für Niedersachsen, Hannover, Germany
[g] Deutsche BKK, Wolfsburg, Germany
[h] Integrated Municipality Schöppenstedt, Schöppenstedt, Germany
[i] County of Wolfenbüttel, Wolfenbüttel, Germany

Abstract

We designed, constructed, and evaluated a mobile medical care vehicle called "Rollende Arztpraxis" (rolling medical practice, RMP) that delivers the full medical care of a general practitioner to increase medical care supply in rural areas. Six communities have been identified, where the RMP has been visited 501 times in 14 months. Two different schedules of stops and treatment times have been tested. We show that the RMP treated mainly elderly and multimorbid patients. An accompanying study showed high acceptance and satisfaction of treated patients and treating doctors. An economic evaluation of three different utilization models with three treatment modes each showed no financial sustainability. We show that ambulatory care in rural areas can be complemented by a mobile care unit, if legal and financial barriers can be overcome.

Keywords:
Rural Health Services; Mobile Health Vans.

Introduction

The German healthcare system is threatened by a lack of doctors [1]. Similarly affected are all communities in northern Germany regarded in the context of this paper. Respective estimations in [2] with a narrower regional scope show significantly reduced medical supply until 2020. Combined with a growing number of elderly, and therefore, more chronically diseased and multimorbid patients, a rising need for medical supply can be assumed. The main objective of the project is to study if ambulatory care in rural areas can be complemented by a mobile care unit called "Rollende Arztpraxis" (rolling medical practice, RMP).

Methods

We investigated technical feasability over 14 months. An accompanying study is split up into four work packages addressing usage figures, satisfaction of patients, satisfaction of doctors and economic sustainability.

Results

The RMP is a commercial utility van with a custom box and a completely adapted interior for the RMPs' primary use case of ambulatory care. Six low-population-communities (A-F) have been selected as stop and treatment places for the RMP. 501 treatments were made. Patients mainly reached the RMP on foot and by bike. Nearly all patients (n=36) surveyed were "very satisfied" or "rather satisfied" with the RMP as a whole. Resident GPs surveyed believed there was an improvement in medical care. Some GPs reacted rather angrily and found the RMP irresponsible or a solely political instrument. According to the mobile doctors, treatment was without any problems. The RMP contains all technical equipment for the treatment of a broad spectrum of patients; many immobile and some disabled patients have been treated without problems. Since no medical assistant was present at the RMP, all organizational work had to be done by the doctors themselves. This resulted in a maximum of 10 to 12 patients a day.

Conclusion

Our technical feasability evaluation showed that full ranged mobile medical care in rural areas is possible if financially subventioned.

References

[1] T. Kopetsch, The German healthcare system is running out of doctors, German Medical Association, Berlin, 2010.
[2] Forecast of physician numbers for the year 2020, Association of Statutory Health Insurance Physicians Lower Saxony, Hannover, 2008.

Address for correspondence

Jonas Schwartze, M.Sc.; Peter L. Reichertz Institute for Medical Informatics; Mühlenpfordtstraße 23, D-38106 Braunschweig;
T +49 (0)531 391-9513; Jonas.Schwartze@plri.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-910

Online Communication and Chronic Obstructive Pulmonary Disease (COPD)

Ina Koldkjær Sølling [a], Per Carøe [a], Kurt Lindgren [b] Kirsten Siggaard Mathiesen [a]

[a] *Department of Nursing, University College of Northern Denmark, Denmark*
[b] *Department of Technology and Business, University College of Northern Denmark, Denmark*

Abstract

This is an explorative and qualitative study that examines a municipal's rehabilitation program "Online Viva" (2014–2015). The questions are whether "Online Viva" improves the citizens' participatory health and whether it prevents exacerbation of COPD. "Online Viva" includes respiratory training and district nursing consultancy for elderly people with COPD. The district nurses' presence in the citizens' homes is replaced by online communication. The study involves 9 citizens and 5 health professionals. Preliminary results show that online consultations and training prevent anxiety, exacerbation, and support the citizens' management of COPD. The citizens find that the "Online Viva" reduces their need for hospitalization, and this is confirmed by the district nurses. Furthermore, the citizens find that their specific needs are fulfilled, making them comfortable in managing their COPD. It is emphasized by the citizens and the district nurses that they must have been in personal contact with each other before they meet online. Furthermore, it must be possible to visit the citizens in their private homes in addition to an online contact, if needed.

Keywords:

Patient centered care, aged care, participatory health, COPD, preventive medicine

Introduction

In 2013, Slagelse Municipality implemented an online rehabilitation program, "Online Viva" with respiratory training and district nursing consultancy for elderly people (aged 61-90) with COPD. In September 2014, 20 citizens with moderate (50%) to severe (15%) reduced lung function were included in "Online Viva", the objective of which was to support the citizens' participatory health and to prevent COPD exacerbation . The intention was to include 30 citizens with COPD or another chronic disease. The demographic development results in more elderly people living longer with chronic diseases, as well as lack of financial resources in the health care sector of the municipalities. There is an assumption that welfare technology can provide cost-effective solutions to the benefit of future citizens despite the strained economy. The technology used in "Online Viva" is based on an online encrypted connection with synchronous video communication. The citizen and the district nurse are able to communicate via the internet at a scheduled time.

Methods

This project is an explorative and qualitative study including 9 citizens (5 women and 4 men aged 62 – 82), 3 district nurses, the supervisor of the project, and the head of the department. Face-to-face semi-structured interviews were conducted with the citizens, the supervisor, and the head of the department. Furthermore, we conducted a focus group interview of the 3 district nurses who provided an audio-reflection about the online consultancy with the citizens (2, 3). In relation to the technology, we chose a socio–technical approach (4). Additionally, we will discus the results with healthy elderly of the DanAge Association (NGO) .to maintain a citizen perspective. The study was approved by the Danish Study of Ethics in Science and Data Protection.

Results

Preliminary results show that online consultations and training prevent anxiety and COPD exacerbation resulting in improved COPD management. . As per citizens' opinion, the "Online Viva" reduces their anxiety and need for hospitalization; confirmed by their district nurses. Additionally, the citizens find that their specific needs are fulfilled making them comfortable in managing COPD. It is emphasized by the citizens and the district nurses that a personal relationship must be established before meeting online. Furthermore, the nurse should be able to visit the citizens in their private homes in addition to an online contact, if needed. The professional knowledge at specialist level and practical experience with technology of the district nurses are necessary requirements for obtaining a confident online communication with the citizens. Furthermore, the district nurse should be able to make decisions, act appropriately, and ensure that the sound and picture are of good quality during online meeting.

Conclusion

The preliminary results indicate that a focused online consultation with respiratory training improves the citizens' participatory health. The consultation and training support the citizens' management of COPD and prevent anxiety, COPD exacerbation, and hospitalization. The district nurse should be well qualified in academic and technological skills and should be able to work at a specialist level.

References

[1] Kvale S, Brinkmann Interviews learning the craft of qualitative research. Sage 2014

[2] Duke J Joining the dots: Piloting the work day as a data collection tool Issues in Educational Research 2012, (22): 111-126

[3] Berg M, Arts J, van der Lei J ICT in health care; sociotechnical approaches. Methods Inf. Med 2003: 42(4) : 297-301

Address for correspondence

Ina Koldkjær Sølling
UCN, Denmark. Mail: iks@ucn.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-911

Development and practice of ISMS at a Radiotherapy Hospital by using IHE Integration profiles

Yutaka Ando[a], Yuki Yoshida[b], Masami Mukai[b], Yasuo Okuda[b], Tadashi Kamada[a], Shuichi Moriguchi[c]

[a] *Research Center for Charged Particle Therapy, National Institute of Radiological Sciences, Chiba, Japan*
[b] *Medical Informatics Section, National Institute of Radiological Sciences,* [c] *MPO Co., Ltd., Yokohama, Japan*

Abstract

Our hospital is specialized for radiation therapy and has many information devices. Various job categories are working. When we implemented an EMR, we aimed to enforce ISMS by using IHE profiles. To solve the already existed system-related problems, we selected and use some profiles (EUA, PSA, ATNA and PAM). After implementation, we audited and then some findings were pointed out. These findings are being settled by the PDCA cycle. We also found that appropriate IHE profiles were effective in the building of ISMS.

Keywords:

ISMS, IHE Integrating the Healthcare Enterprise, PDCA

Introduction

Our hospital is specialized for a radiation therapy with a Carbon ion. We have 100 beds for inpatient, 70-110 outpatients/day and two departments (Radiology and Dentistry). There are many special treatment devices and information terminals in our hospital. And our staff consists of many occupational categories such as physicians, nurses, technicians, operators and clerks. Complexity and diversity of our information system may hold some problems. These problems are: (1) single sign on, (2) synchronizing patient, (3) time consistency and audit log and (4) co-operation among information systems.

The promotion of Information Security Management System (ISMS) is a key issue. We use some integration profiles of Integrating the Healthcare Enterprise (IHE) to improve the security of information system and to solve these problems. We evaluate the IHE profiles in the aspect of security management.

Materials and Methods

We implemented some IHE profiles to solve some clinical problems mentioned above. Methods we used were Enterprise User Authentication (EUA) for single sign in, Patient Synchronized Applications (PSA) for co-operation of same patient, Audit Trail and Node Authentication (ATNA) for audit log, and Patient Administration Management (PAM) for patient information reconciliation. These EUA/PSA functions are protected from misselection of patient and forgotten logout.

In 2006 our hospital had prepared to strengthen information security and we detected the personal information and analyzed the system's risk. Also we had defined regulations concerning handling and protection of personal information.

From 2007 we have continued the audit about our privacy policy and security policy according to what ISMS established. We practiced this audit once per year. The audit result proceeds to the management review of the hospital director. The PDCA cycle consists of Plan, Do, Check and Act. We then executed this cycle (audit, analysis of problems, action plan, manual revise and check list) in a year. These processes are shown in Figure 1.

Results

Audit findings are as follows: (1) An unimproved rule about cancellation of a user account and abolition, (2) A connection of an internal network and the internet, (3) An inadequate periodic supervision and management about practical use of a USB memory, (4) An incomplete key management of the outpatient's examination room, server room, rocker and security chain of the PC body, (5) Insufficiency in the password management and (6) An unclear responsibility of procedure of a wireless LAN and operations management.

The PDCA cycles decreased these problems. Finally at the promotion of ISMS, we feel undoubtedly that these problems have to be solved. They are (1) Improvement of motivation for ISMS implementation, (2) Complexity of risk analysis in a medical field and (3) Difficulty of evaluating the weight of confidentiality, integrity and availability. Hospital staffs were familiar with the PDCA cycle and did not offer resistance to use PDCA cycle. We think that the promotion of ISMS depends on a deep understanding and strong leadership.

Figure 1 – PDCA cycle in our hospital

Conclusion

We implemented EMR and planned to improve information security using IHE profiles. We utilized PDCA cycle to raise information security level. In addition, we believe that when we select the suitable IHE profiles, we can achieve the reinforcement of the secure information system at the hospital information systems.

Address for correspondence

Yutaka Ando MD: ando_y@nirs.go.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-912

Optimizing Safety, Fidelity and Usability of an Intelligent Clinical Support Tool (ICST) For Acute Hospital Care: an Australian Case Study Using a Multi-Method Delphi Process

Mari Botti[a], Bernice Redley [a], Lemai Nguyen[b], Kimberley Coleman [a], Nilmini Wickramasinghe[c]

[a]*Epworth HealthCare/Deakin University Centre for Clinical Nursing Research*
[b]*Deakin University School of Information and Business Analytics*
[c]*Epworth HealthCare/RMIT University Health Information Management*

Abstract

This research focuses on a major health priority for Australia by addressing existing gaps in the implementation of nursing informatics solutions in healthcare. It serves to inform the successful deployment of IT solutions designed to support patient-centered, frontline acute healthcare delivery by multidisciplinary care teams. The outcomes can guide future evaluations of the contribution of IT solutions to the efficiency, safety and quality of care delivery in acute hospital settings.

Keywords:

Nursing informatics; patient safety; user-centred design; acute healthcare delivery

Introduction

Technology innovations in medical informatics have transformed many areas within healthcare delivery; however, poor uptake of technology remains a widespread problem. Technology solutions to support clinical work are frequently engineered from administration requirements rather than the needs of front-end clinicians; making them confusing to users, incompatible with existing clinical workflows, increasing workloads and creating potential risks to patients [1-4]. As a result, clinicians often either abandon these systems or are highly resistant to using them.

This study examines the benefits of adopting a user-centred design approach in developing and deploying an intelligent clinical support tool (ICST) in acute clinical settings in public and private hospitals in Melbourne, Australia. The methodological approach designed to address this issue is presented as an appropriate, systematic strategy that mitigates safety risks and enhances the likelihood of user adoption and acceptance.

Methods

The study adopts a multi-method approach using qualitative and quantitative techniques. A key unique feature of the research is the Delphi process used to elicit multi-disciplinary responses and interactions with the system. Critical phases of the methodology include: **Phase 1** - Establish fidelity: Interviews with key multidisciplinary users to explore their specific information needs, interactions, work flow, data inputs and outputs associated with patient care delivery. **Phase 2** – Evaluate fidelity: A Delphi process to examine multidisciplinary agreement on system design for patient safety will include up to six improvement cycles and one consensus cycle over twelve months. Each cycle will incorporate discipline specific data into the ICST and observe participants using the technology in simulated clinical scenarios. 'Think- aloud' [5] techniques will be used to elicit participants' experiences when interacting with the system and interface. **Phase 3** – Reflections on fidelity and safety: Brief interviews with participants at the end of each Delphi cycle. An online survey of system usability and fidelity will be collected one week after each cycle.

Results

The continuous feedback and cyclic improvement of the system provide the possibility to design, develop and deploy a sophisticated clinical informatics solution that caters to the real workflow needs of multidisciplinary clinicians. This serves to ensure superior levels of usability, safety and fidelity for clinical practice, as well as reduces the likelihood of resistance and lack of use of the technology solution. At this stage of this research in progress we are in Phase 1 to establish fidelity. Interviews with key users have enabled us to identify the different user needs and ensure that the solution has the capability to address these respective needs.

Conclusions

To date we have identified commonalities and discipline specific needs to inform system design. This research has far reaching implications for the design, development and deployment of sustainable informatics solutions that ensure the safety of patients while streamlining and improving clinical practice through technological innovation.

References

[1] Urquhart C, Currell R. Reviewing the evidence on nursing record systems. Health Informatics Journal. 2005;11(1):33-44.

[2] Sittig DF, Singh H. Electronic Health Records and National Patient-Safety Goals. New England Journal of Medicine. 2012;367(19):1854-60.

[3] Aarts J. Towards Safe Electronic Health Records: a socio-technical perspective and the need for incident reporting. Health Policy and Technology. 2012;1(1):8-15.

[4] Eliza MM, Peter AB. Impact of Health Informatics Implementation on Clinical Workflow: A Review. Lecture Notes in Engineering and Computer Science. 2012(1):1352.

[5] Cotton D, Gresty K. Reflecting on the think-aloud method for evaluating e-learning. British Journal of Educational Technology. 2006;37(1):45-54.

Address for correspondence

Deakin-Epworth Centre for Clinical Nursing Research,185-187 Hoddle St, Richmond Victoria, AUSTRALIA 3121

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-913

An Integrated Workflow For Secondary Use of Patient Data for Clinical Research

Guillaume Bouzillé[a,b,c], Emmanuelle Sylvestre[a,b,c], Boris Campillo-Gimenez[a,b], Eric Renault[a,b], Thibault Ledieu[a,b], Denis Delamarre[a,b], Marc Cuggia[a,b,c]

[a] *INSERM, U1099, Rennes, F-35000, France*
[b] *Université de Rennes 1, LTSI, Rennes, F-35000, France*
[c] *CHU Rennes, CIC Inserm 1414, Rennes, F-35000, France*

Abstract

This work proposes an integrated workflow for secondary use of medical data to serve feasibility studies, and the prescreening and monitoring of research studies. All research issues are initially addressed by the Clinical Research Office through a research portal and subsequently redirected to relevant experts in the determined field of concentration. For secondary use of data, the workflow is then based on the clinical data warehouse of the hospital. A datamart with potentially eligible research candidates is constructed. Datamarts can either produce aggregated data, de-identified data, or identified data, according to the kind of study being treated. In conclusion, integrating the secondary use of data process into a general research workflow allows visibility of information technologies and improves the accessability of clinical data.

Keywords:

Biomedical research; Secondary use of data; Confidentiality

Introduction

With the development of electronic health records (EHRs) secondary use of patient data is becoming a critical issue for medical research. Hospitals are currently deploying information technologies (IT) and tools intended to facilitate easy access to clinical data for research purposes.

In this paper, we present an integrated workflow that aims to manage and to monitor the global process of reusing clinical data for research purposes. This workflow system addresses three main scenarios of clinical research: feasibility study, pre-screening or cohort building, and monitoring. It is currently used and tested by the Clinical Research Office (CRO) of the Academic Medical Centre of Rennes (AMC).

Methods

The workflow is based on two main applications: a research portal dedicated to the research activities in the Academic Medical Centre of Rennes and EHOP, the clinical data warehouse (CDW).

The portal and the information retrieval interface of the CDW were developed as web applications employing the following technologies: Angular.js framework, Java technologies (portal) or php (CDW), Oracle database, and Oracle text functionalities. The system was integrated in the Hospital Information System (HIS).

Results

The goal of the research portal was to give to physicians a platform for any research purposes such as methodology, data analysis, data extraction or the global management of a research project. Each request was primarily treated by the CRO and then assigned to one or more resource experts within each specialized field. Any ticket concerning reuse of patient data were addressed to the experts of the data-mining department of the hospital.

Prior to the granting of physician access to patient data, several requirements were fulfilled. The expert must select from aggregated, de-identified, or identified individual data, the type of access required for the study. Ethical and legal procedures to be followed by the physician were contingent on study-type. The research portal aimed to be a support to assist the physician for theses procedures and to monitor activity.

The next step was to identify the characteristics of the studied population in order to fulfil all the necessary elements to request the CDW for candidate patients.

Access to the entire dataset stored in the clinical data warehouse was limited to experts of the data-mining department. They were responsible for extracting a subset of potential patients (except those who refused to have their data used for research purposes) of interest for the study. Data were then replicated in a datamart and made available for physicians for a limited time through a link available in the study's ticket. Physicians were thus able to mine candidate patients' data to check their eligibility for the study by using Datamart information retrieval tools.

Conclusion

Developing relevant tools to support the reuse of patient data and especially to manage every step of the workflow appears to be crucial. Integrating the secondary use of data process in a general research workflow gives visibility to information technologies and makes access to clinical data easier. Moreover, it allows traceability of access to data and monitoring of activity of secondary use of patient data.

Address for correspondence

guillaume.bouzille@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-914

Screening Mammography Efficacy: A Comparison Between Screen-Film, Computed Radiography and Digital Mammography in Taiwan

Dr. Abdallah Ahmed Elbakkoush[a, 1], Suleman Atique [a], DR.I-Jen Chiang [a, 1]

[a] *Graduate Institute of Bio-medical Informatics, Taipei Medical University, Taiwan*

Abstract

Breast cancer screening has been proven effective in Western countries, and our National Health Bureau also started to offer free screening mammography for women aged between 50 and 69, since July 2004. This study aims to compare the efficacy between distinctive mammography screening modalities. Prompting the assessment of digital screening in the radiological sciences, we provide insight into the practical informatics application of such tools.

Keywords:

Computed Radiography; Digital Mammography, Screen Film, Screening method

Methods

Between July 2004 and December 2010, 23513 eligible women underwent screening mammography. 6804 of these received screen-film mammography, 1610 received computed-radiography mammography, and 15099 received digital mammography. The carcinoma-in-situ detection rate, early cancer detection rate, and the overall cancer detection rate of each modality were calculated. Chi-square test was used to determine the statistical significance of performance differences between these modalities. Different modalities, which are derived from informatics applications in the radiological sciences, are compared in this paper.

Results

Screen-film mammography's callback rate was 12.23%, cancer detection rate was 5.29%, and Ductal Carcinoma In Situ (DCIS) detection rate was 1.76%. Computed Radiography's callback rate was 12.67%, with a cancer detection rate of 6.21%, and DCIS detection rate of 1.24%. Full-field digital mammography (FFDM) had the highest callback rate of 9.21%, but also had the highest rate of cancer detection rate and DCIS detection rate, 9.21% and 3.58%, respectively.

FFDM had better cancer detection rate and DCIS detection rate compared to Screen-Film Mammography (SFM) and Computed Radiography (CR), however, this only demonstrated statistical significance when compared to SFM. Although FFDM had better cancer and DCIS detection rates, it was also associated with a higher callback rate. Results show that the employment of digital modalities can enhance the understanding and application of mammography tools in the radiological sciences.

Discussion

The cancer detection rate for Digital Mammography was higher than both SFM and CR Mammography in the screening of breast cancer and ductal carcinoma in situ.

Unlike prior reports, our results show that the incidence rate for Taiwan women is not low. CR mammography had a lower DCIS detection rate, compared with DR, in Taiwan women dense breast. If we rely on CR, we may fail to identify many DCIS cases, which is an important consideration. As the price of DM gradually decreases, by using DM, facilities may save on storage space, retrieval and processing time, and promote a more comfortable environment for patients. In order to improve the efficiency, quality, and safety of patient care, there is need to further explore such tools in the field of radiology.

Conclusion

Full-field digital mammography has obvious performance advantages over screen-film and computed radiography mammography. This is important, especially in Asian populations where the prevalence of dense breasts is increased. Taiwanese breast cancer incidence is gradually increasing and becoming comparable to that of Western countries. Such factors, in addition to Taiwan's unique medicolegal ecosystem, result in relatively high callback rates.

Address for correspondence

Graduate Institute of Bio-medical Informatics, Taipei Medical University, Taiwan, Wuxing Steet 250, Taipei 110, ROC(Taiwan)

[1]Correspondence Authors:

DR.I-Jen Chiang E-mail

ijchiang@tmu.edu.tw

Dr.Abdallah Ahmed Elbakkoush

mailto:alosta20102001@yahoo.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-915

Evaluation of a Cyber Security System for Hospital Network

Mohammad A. Faysel[a]

[a] *Medical Informatics Program, SUNY Downstate Medical Center, Brooklyn, NY, USA*

Abstract

Most of the cyber security systems use simulated data in evaluating their detection capabilities. The proposed cyber security system utilizes real hospital network connections. It uses a probabilistic data mining algorithm to detect anomalous events and takes appropriate response in real-time. On an evaluation using real-world hospital network data consisting of incoming network connections collected for a 24-hour period, the proposed system detected 15 unusual connections which were undetected by a commercial intrusion prevention system for the same network connections. Evaluation of the proposed system shows a potential to secure protected patient health information on a hospital network.

Keywords:

Cyber security; Hospital network; Health information system.

Introduction

Due to recent advancement in health information technologies and mandatory adoption of electronic health records, dependency on the Internet is unavoidable for accessing patient health information [1]. Hospitals are relying on commercially available network security tools such as firewalls, intrusion detection system (IDS), and intrusion prevention system (IPS) to secure information network. These tools have already been proven to be inadequate to defend real-time cyber security incidences [2]. The proposed system utilizes packet header information from incoming network connections and applies a probabilistic data mining technique to detect cyber security incidences in real-time. When an anomalous network connection is detected, it also generates real-time alert messages to the hospital network administrator.

Methods

The cyber security system consists of three modules- a network connection capture module, a detection module, and a response module. The network connection capture module collects incoming IP packet headers generated by perimeter firewalls. It uses Netflow [3], a protocol developed by Cisco Inc., to capture the following fields related to a network connection: source IP address, destination IP address, source port, destination port, TCP Flags, protocol, and time. The detection module utilizes these packet information and applies a probabilistic data mining algorithm to calculate a normality score for each connection [4]. The algorithm uses values of these fields as inputs and compares them to the most recent week's network connections. Then, it calculates a probability score for each of the field based on the presence of the values in the past week's network connections; and assigns a normality score to each incoming connection as sum of these scores. The algorithm then compares this normality score to a threshold (the lowest normality score in the past week's attack free network connections). If a connection score exceeds the threshold, it labels it as an incidence and invokes the response module. It also creates an attack signature based on the incidence. The response module generates alert notification to notify the administrator in real-time for further action.

Results

The system was implemented using MySQL 5.1 database environment. It was evaluated on network connections collected for one day period from a real hospital network. The evaluation dataset consisted of 3,850 unfiltered incoming network connections. The system accurately detected 15 unusual network connections during the evaluation which were visually inspected and verified as anomalous. The results were compared to the system log of a commercial IPS that was protecting the hospital network for the same time period. The IPS was unable to detect those 15 unusual connections which were detected by the proposed system.

Conclusion

The evaluation was done using real network connection from one hospital. The author plans to further evaluate the system on multiple hospital networks in real-time. Based on the evaluation results, the proposed system outperformed the commercial IPS tool in detecting unusual incidences. In addition, it shows a potential to defend cyber security incidences on hospital network.

References

[1] Appari A, Johnson ME. Information security and privacy in healthcare: current state of research. IJIEM 2010; 6(4): 279-314.

[2] Faysel MA, Haque SS. Towards cyber defense: research in intrusion detection and intrusion prevention systems. IJCSNS 2010; 7:316-325.

[3] Nickless B. Combining cisco netflow exports with relational database technology for usage statistics, intrusion detection, and network forensics. Proceedings of the 14th USENIX Conference on System Administration; 3-8 December 2000; New Orleans, Louisiana. 2000. pp. 285-290.

[4] Faysel MA, Haque SS. A novel algorithm to defend cyber attacks in hospital information networks. IJCIHI 2013;6:29-31.

Address for correspondence

Mohammad A. Faysel, Ph.D.

Assistant Professor

Medical Informatics Program

450 Clarkson Avenue, Mail Stop 105, Brooklyn, NY 11203, USA

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-916

Design and development of an EMR for Ebola Treatment Centers in Sierra Leone using OpenMRS

Darius Jazayeri [a,g], Shefali Oza [b], Glauber Ramos [a], Hamish Fraser [c,e], Jonathan M. Teich [d,e], Andrew S. Kanter [f], Ellen Ball [g]

[a] *ThoughtWorks Uganda Ltd., Kampala, Uganda.* [b] *London School of Hygiene and Tropical Medicine, London, UK*
[c] *University of Leeds, Leeds, UK.* [d] *Brigham and Women's Hospital, Boston, USA.* [e] *Harvard University, Boston, USA*
[f] *Columbia University, NY, USA.* [g] *Partners In Health, Boston, USA*

Abstract

Ebola treatment presents unique challenges for medical records because strict infection control requirements rule out most conventional record-keeping systems. We used the OpenMRS platform to rapidly develop an EMR system for the recently opened Kerry Town, Sierra Leone Ebola Treatment Centre. This system addresses the need for recording patient data and communicating it between the infectious and non-infectious zones, and is specifically designed for maximum usability by staff wearing cumbersome protective equipment. This platform is interoperable with other key eHealth systems in the country, and is extensible to other sites and diseases.

Keywords:

EMR; Ebola; OpenMRS.

Introduction

The 2014 West African Ebola outbreak is unprecedented in its scale and spread, with nearly 20,000 reported cases by December 2014. Ebola Treatment Centres (ETCs) are specialized facilities that must provide efficient care for suspected and confirmed Ebola patients while minimizing risk of cross infection to staff and other patients. This presents several challenges, including the difficulty of working in personal protective equipment (PPE) due to heat and low visibility/dexterity, and the inability to move material from highly infectious patient areas (Red Zone) to low-risk areas (Green Zone). Conventional paper-based data collection is difficult for these reasons. ETCs are therefore often limited in their ability to collect data for care, surveillance, and research.

Important ETC information tasks include medication ordering and monitoring; initial, daily and per-shift patient assessments; lab test ordering and results; and fluid management. These require data entry and review within the Red Zone, as well as communication and coordination between zones. Here, we describe the rapid development of an innovative Ebola EMR system based on the OpenMRS open-source EMR platform [1]. This system is designed specifically to address the usability, workflow and communication problems noted above.

Materials and Methods

We built an Ebola module for the OpenMRS 2.1 platform for the Kerry Town ETC, run by Save the Children with UK government funding. OpenMRS is a widely implemented flexible open-source, modular EMR platform currently used in over 50 low and middle-income countries. We used Agile software development methodology. Detailed gathering of requirements was done with on-site staff, and domain experts provided design input. We rapidly iterated on designs using wireframes and live prototypes.

Results

The first release (Feb. 2015) included patient registration, limited clinical data entry and a comprehensive drug ordering/monitoring system for the Red Zone and pharmacy. We optimized the tablet-based user interface for PPE use with large text and touch targets, minimal typing, a high-contrast colour scheme, and easy selection of items like commonly used drugs. We used core OpenMRS concepts, data structures and functional modules, and added Ebola-specific core concepts to simplify decision support and reporting across ETC sites. The next release included modules for vital signs, clinical assessment and fluid balance. Later steps include coordination with case finding, contact tracing and lab testing.

Conclusions

Careful user design on a flexible platform can rapidly yield EMRs that are well suited to address the unique data challenges in Ebola treatment centers. Our system can be adapted for other ETCs and for future health crises.

Acknowledgments

We acknowledge the support of the OpenMRS community, ThoughtWorks, Inc, and Save the Children International.

References

[1] Mamlin BW, et al. Cooking up an open source EMR for developing countries: OpenMRS – A recipe for successful collaboration. Proc AMIA Symp 2006: 529-533

Address for correspondence

Darius Jazayeri: djazayer@thoughtworks.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-917

The Impact of an Electronic Medical Record on Repeat Laboratory Test Ordering Across Four Australian Hospitals

Andrew Georgiou[a], Elia Vecellio[a], Ling Li[a], Johanna I. Westbrook[a]

[a] *Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Australia*

Abstract

In this study we examined the impact of an Electronic Medical Record (EMR) on repeat test rates (i.e., the same test ordered within a specified window of time) for a commonly ordered set of laboratory tests; Electrolytes, Urea, Creatinine [EUC], Full Blood Counts [FBC] and Liver Function Tests [LFT]. The results point to the potential that timely, evidence-based electronic decision support features can have on the efficiency and effectiveness of the pathology laboratory process and its contribution to quality patient care.

Keywords:

Computerised provider order entry, Evaluation studies, Hospital information systems, Laboratories

Introduction

Electronic Medical Record (EMR) systems can include Computerised Provider Order Entry (CPOE) functionality, and provide for hospital-wide integration of electronic clinical and patient databases. In this study we examined the impact of an EMR on clinician laboratory test ordering patterns by comparing paper and EMR repeat test rates (i.e., the same test ordered within a specified window of time) for a commonly ordered set of laboratory tests (Electrolytes, Urea, Creatinine [EUC], Full Blood Counts [FBC] and Liver Function Tests [LFT].

Material and Methods

Laboratory data were extracted for the period of Aug/Sep 2008-2011 across four hospitals serviced by a single pathology laboratory service. The EMR was initially based on the Cerner PowerChart system Version 2007.16, and upgraded in May 2011 to Version 2010.02.16. The Laboratory Information System (LIS) was the Integrated Software Solutions (ISS) Omnilab v9.4.2 SR10 system. The electronic ordering system presented clinicians with information about previous test orders and results, and alerted clinicians when a repeat test was ordered within 24-hours of the previous test. Ethics approval was provided by the relevant Local Health District Human Research Ethics Committee (HREC: Project No. 11/146) The project was funded by an Australian Government Department of Health: Quality Use of Pathology Program grant.

Results

Overall, for all the data collected, the percentage of repeat tests (which were ordered using either the EMR or paper) was 77.2% (136644/177096) for EUC, 75.1% (126070/167791) for FBC and 68.9% (68953/100019) for LFT. The results show significant ($p<.0001$) reductions in repeat test rates within 1 hour following the introduction of the EMR. Tests for EUC decreased by 0.2% for EUC, 0.1% for FBC and 0.2% for LFT. For repeat tests within 12 hours, there were significant ($p<.0001$) decreases in EMR repeat orders of 1.1% for EUC and 1.2% for FBC. There were also significant ($p<.0001$) decreases of 1.8% for both EUC and FBC repeat tests within 24 hours. For LFT the EMR repeat test order rate increased significantly ($P<.0001$) by 1.8% within 24 hours and 4% with 36 hours. These results indicate a greater concordance with best practice guidelines.

Conclusion

Our results confirm previous evidence about the impact of the EMR on the rate of redundant or unnecessary tests [1,2]. Such improvements are connected to EMR's ability to provide clinicians with an overview of existing test orders for each patient along with a visual aid alerting them to repeat tests undertaken within a "too-short" re-test interval. Taken together these features can enhance clinicians' ability to monitor and regulate their test ordering practices.

References

[1] Georgiou A, et al. Troponin testing in the emergency department: a longitudinal study to assess the impact and sustainability of decision support strategies. Journal of Clinical Pathology. 2012;65(6):546-50.

[2] Bates D, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. American Journal of Medicine. 1999;106(2):144-50.

Address for correspondence

A/Prof Andrew Georgiou (andrew.georgiou@mq.edu.au)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-918

Implementation of Data Driven Heart Rate and Respiratory Rate Parameters on a Pediatric Acute Care Unit

Veena Goel[a,b], Sarah Poole[b], Alaina Kipps[a,b], Jonathan Palma[a,b], Terry Platchek[a,b], Natalie Pageler[a,b], Christopher Longhurst[a,b], Paul Sharek[a,b]

[a] *Stanford Children's Health, Palo Alto, CA*
[b] *Stanford University School of Medicine, Palo Alto, CA*

Abstract

The majority of hospital physiologic monitor alarms are not clinically actionable and contribute to alarm fatigue. In 2014, The Joint Commission declared alarm safety as a National Patient Safety Goal and urged prompt action by hospitals to mitigate the issue [1]. It has been demonstrated that vital signs in hospitalized children are quite different from currently accepted reference ranges [2]. Implementation of data-driven, age stratified vital sign parameters (Table 1) for alarms in this patient population could reduce alarm frequency.

Keywords:
Clinical Alarms; Vital Signs

Methods

This is a prospective study of alarm frequency after bedside monitor implementation of age stratified data-driven heart rate (HR) and respiratory rate (RR) parameters on a cardiac medical/surgical step-down unit at a quarternary-care children's hospital. The data-driven parameters were derived using nurse-documented HR and RR vital signs (62,508 unique measurements) of all non-ICU hospitalized patients <18 yrs during calendar year 2013 at Lucile Packard Children's Hospital Stanford (LPCH). Alarm frequency data was analyzed over 28 day periods before and after the intervention on 10/27/14. HR and RR alarm data were obtained from the LPCH clinical data warehouse, which stores real-time central monitor system data.

Results

In the pre-intervention study period, there were a total 7,028 HR alarms and 8,400 RR alarms, yielding an average of 14.1 HR alarms/patient day and 16.9 RR alarms/patient day. In the post-intervention period, total HR alarms decreased to 4,844 (9.6/patient day), but RR alarms increased to 20,860 (41.3/patient day).

Discussion

While the frequency of HR alarms decreased following implementation of age-stratified, data-driven vital sign parameters, the frequency of RR alarms markedly increased, particularly for low RR values (*Figure 1*). Factors contributing to this discrepancy might include: (1) inaccurate creation of data-driven parameters because nurse-charted values were used to calculate them rather than raw monitor data, or because representative patients who spent time in the ICU were excluded from the analysis; and (2) it is possible that pre-intervention monitor parameter settings did not represent pre-intervention reference ranges and so when default monitor settings were updated as part of the intervention, the alarm rate increased as a result of this discrepancy.

Conclusion

Our results suggest that adoption of data-driven values for HR alarm parameters will decrease the frequency of HR alarms in a children's hospital. However, further work to understand the increase in RR alarms is necessary, highlighting the challenges in creation and implementation of data-driven vital sign parameters to reduce alarm fatigue.

Table 1– Data-driven HR and RR limits, by age

Age	Low HR	High HR	Low RR	High RR
<1month	<115	>170	<30	>60
1- <6mo	<105	>170	<20	>55
6mo- <1yr	<100	>165	<20	>45
1- <2yr	<90	>165	<20	>45
2- <3yr	<85	>155	<18	>40
3- <5yr	<75	>155	<16	>35
5 -<9yr	<70	>140	<16	>30
9- <12yr	<65	>130	<14	>30
12- <15yr	<60	>125	<14	>30
≥15yr	<60	>115	<13	>25

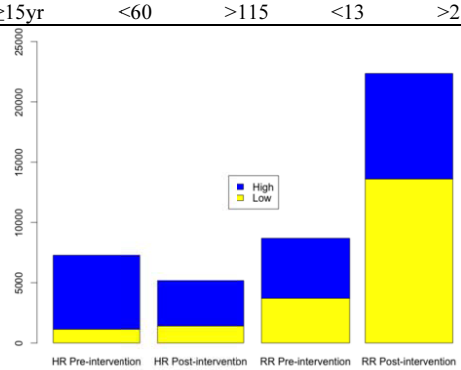


Figure 1– Pre- and post-intervention HR & RR alarms.

References

[1] The Joint Commission. 2014 National Patient Safety Goal on Clinical Alarm Safety. 2013.
[2] Bonafide CP, Brady PW *et al.* Development of Heart and Respiratory Rate Percentile Curves for Hospitalized Children. *Pediatrics.* Apr 2013;131(4):e1150-1157.

Address for correspondence
Veena Goel vgoel@stanfordchildrens.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-919

Application of Barcoding to Reduce Error of Patient Identification and to Increase Patient's Information Confidentiality of Test Tube Labelling in a Psychiatric Teaching Hospital

Hsiu-Chu Liu[a & b], Hsing Li[b], Hsin-Fei Chang[b], Mei-Rou Lu[b], Feng-Chuan Chen[c]

[a] *EMBA Student, National Pingtung University of Science and Technology, Taiwan*
[b] *Department of Nursing, Kaohsiung Municipal Kai-Suan Psychiatric Hospital, Taiwan*
[c] *Informatics office of Planning Department, Kaohsiung Municipal Kai-Suan Psychiatric Hospital, Taiwan*

Abstract

Learning from the experience of another medical center in Taiwan, Kaohsiung Municipal Kai-Suan Psychiatric Hospital has changed the nursing informatics system step by step in the past year and a half . We considered ethics in the original idea of implementing barcodes on the test tube labels to process the identification of the psychiatric patients. The main aims of this project are to maintain the confidential information and to transport the sample effectively. The primary nurses had been using different work sheets for this project to ensure the acceptance of the new barcode system. In the past two years the errors in the blood testing process were as high as 11,000 in 14,000 events per year, resulting in wastage of resources. The actions taken by the nurses and the new barcode system implementation can improve the clinical nursing care quality, safety of the patients, and efficiency, while decreasing the cost due to the human error.

Keywords:

Barcoding; Identification; Confidentiality; Psychiatric Nursing; Inpatients; Health Care Quality; Technology Acceptance

Introduction

Under the supervision of the city health department and accreditation, the psychiatric teaching hospital needed to improve its nursing informatics system. The new barcode system was influenced by psychiatric patients' right. Double identification of the patients is important to ensure that the staff do not make a mistake of performing wrong test on the patients. However, the psychiatric patients could be uncooperative to the examination and may be unwilling to give the staff their identification information.

Materials and Methods

The questionnaire with experts' validity will be a resource to understand patients' protection, technology acceptance, and satisfaction of the nursing staff. The experts' questionnaire will check Cronbach's α reliability and modify the structured questionnaire to the most streamlined version in two 50-bed acute wards with 30 nursing participants.

The before and after the implementatopm comparison will show the extent of improvement by the barcode system. The acceptance of the nursing staff will also be collected by the questionaire, and the corresponding analysis using the SPSS 17.0 software will assess the usefulness and the ease of use of the barcode system.

Results

We expect that this project will 1) decrease the number of errors in the psychiatric patients' blood test processing, and 2) improve the understanding of the factors related to the acceptance of barcode system.

We appreciate Kaohsiung Chang Gung Memorial Hospital for providing the barcode system used in the general settings for target learning. We also like to thank superior management officers for their continuous support.

Reference

[1] Alquraini, H., Alhashem, A. M., Sham, M. A., & Chowdhury R. I., Factors influencing nurses' attitudes towards the use of computerized health information systems in Kuwaiti hospitals. Journal of Advanced Nursing, 57 (4), 2007; pp.375-381.

[2] Hsiao, JL., Chang, HC., & Chen, RF., A Study of Factors Affecting Acceptance of Hospital Information Systems: A Nursing Perspective, Journal of Nursing Research, 19(2), June 2011; pp. 150-160.

[3] McGowan C, Patients' Confidentiality. Critical Care Nurse, 32(5), 2012; pp. 61-65.

[4] Sandler, SG, Langeberg, A, & Dohnalek, L., Barcode technology improves positive patient identification and transfusion safety. Dev Biol (Basel), 2005; 120:pp. 19-24

[5] U.S. Department of Health and Human Services. OCR Privacy Brief. Summary of the HIPAA Privacy Rule. http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf. Accessed July 6, 2012.

Address for correspondence

Ms. Hsiu-Chu Liu
e-mail address: elitedoris@yahoo.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-920

Open Source Software For Patient Data Management In Critical Care

Jacques Massaut[a,b,c], Nicolas Charretk[b,c], Olivia Gayraud[c], Rafael Van Den Bergh[c], Adelin Charles[b], Nathalie Edema[b]

[a] *Intensive Care, Brussels Free University, Brussels, BELGIUM*
[b] *Tabarre Trauma Center, Port Au Prince, HAITI*
[c] *Médecins Sans Frontières OCB*

Abstract and Objective

We have previously developed a Patient Data Management System for Intensive Care based on Open Source Software. The aim of this work was to adapt this software to use in Emergency Departments in low resource environments. The new software includes facilities for utilization of the South African Triage Scale and prediction of mortality based on independent predictive factors derived from data from the Tabarre Emergency Trauma Center in Port au Prince, Haiti.

Keywords:

Critical Care, Patient Data Management Software, Open Source Software, Patient Triage Systems.

Introduction

The system for Intensive Care (IC) was designed using a client-server architecture, running on a PostgreSQL database. It offers functions for medical notes, observations and treatments, a scoring system for classification, and reporting of evolution. The system includes utilities for nursing charts and medication administration. To be used in an Emergency Department (ED), the system had to be adapted, particularly to facilitate triage at admission and to propose an adapted score.

Methods

The software was extended to the use of the South African Triage Scale (SATS)[1] and of the TRISS scoring system using the same software architecture as before. For mortality predictions using TRISS, computations were based on the data published by Boyd *et al*[2]. For the SATS system, we did not find any existing publications reporting mortality prediction. To determine mortality prediction, we decided to use data from patients admitted to the Tabarre Emergency Trauma Center from Médecins Sans Frontières (MSF) in Port au Prince, Haiti, as the modified software was intended to be used there. Data from 4,468 consecutive patients admitted to that institution from 01/01/2013 to 08/03/2014 were used. Age of the patients, motif of admission according to MSF classification, and triage category (Red, Orange, Yellow, Green) according to the SATS system were extracted from the local MSF database and used as candidate parameters for analysis of factors associated with mortality. For the statistical analysis we used the Chi Square and Mann-Whitney tests as appropriate to determine association with mortality, and logistic regression with logit to compute independent odds ratios and predict mortality. A p-value < 0.05 was considered to be significant. For logistic regression we divided patients in two groups at random. The first group was used for the computations of odds ratios and coefficients and the second for validation of the model. The STATA 8 software for Windows was used for the statistical analysis.

Results

Added modules to the software help for triage at admission in the ED, and for data encoding. Software includes procedures to compute mortality predictions based on the SATS and TRISS systems. The model included for the SATS system is based on data from Tabarre. Mortality varied significantly with age of the patients, motif of admission, and triage category. Odds ratios for factors affecting mortality with coefficients of logit estimates for that model are illustrated in Table 1.

Table 1 – Factors affecting mortality

Factors	Odds Ratios	Coef.	Z	P > z
Age > 45	2.62	0.963	2.52	0.012
Red	78.98	4.369	7.54	0.000
Orange	3.67	1.300	2.18	0.029
Non Trauma	6.80	1.916	3.56	0.000
(Constant)		-6.317	-11.58	0.000

Area under the ROC curve is 0.8758 for the first group and 0.8344 for the validation group. In the validation group 98.3% of the admitted cases were correctly classified (cutoff of 0.5). In the studied population, the overall mortality was 1.68 %, the mortalities for the SATS groups were 15.81% in the Red group and 1.39% in the Orange. Mortality for patients older than 45 years was 3.29%, versus 1.33% for younger patients.

Conclusion

Our software was extended for the use in the ED, and includes mortality prediction based on the South African Triage Scale. It should now be tested in ED and IC in low resource environments with extension of the data set used in the model.

References

[1] Twomey M.Wallis LA. Thomson ML.Meyers JE. The South African Triage Scale (adult version) provides valid acuity ratings when used by doctors and enrolled nursing assistants. African J.Emerg Med 2012;2:3-12.

[2] Boyd CR, Tolson MA, Copes WS: Evaluating Trauma Care: The TRISS Method. J.Trauma 1987;27:370-378.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-921

Measuring ICTs Adoption in Health Care Facilities in Uruguay

Cecilia Muxí[a], Juan Gil[a], Elisa Martínez[a], Betania Arispe[a], Rossana Occhiuzzi[a], Juan Lozano[a], Jorge Forcella[a]

[a] *Salud.uy Program, eHealth National Initiative, Montevideo, Uruguay.*

Abstract

The use of Information and Communication Technologies (ICTs) has gained an important role to enhance healthcare systems by increasing access to services, and improving the quality and equity of care. In order to evaluate the stage of ICTs adoption in the Uruguayan healthcare system, the Governmental Salud.uy Program carried out a national survey to establish the baseline for ICTs adoption, as well as define a set of indicators so as to evaluate the progress on this topic. Data was recorded following the OECD methodology.

***Keywords*:**

ICTs adoption; healthcare system; maturity assessment.

Introduction

The Uruguayan health reform (2007) aimed to increase equity in access and focus on transparency, efficacy and efficiency to improve quality of care through the creation of a National Health Insurance System. ICT serve as the tool to contribute to health services to interoperate and to decrease the health care system's high degree of fragmentation. Due to the lack of complete data and the need of a national analytic maturity assessment tool, the Program carried out a survey in order to measure progress in ICT adoption.

Methods

The survey was carried out from March to June 2014. Two scenarios were considered: healthcare professionals and health institutions. Physicians and nurses registered in the Ministry of Health database were considered for the 600 individuals sample, stratified by professional category. Data was collected by means of two structured questionnaires, following the OECD Guide [1] and the Cetic.Br regional adaptation model [2].

Results

***Technology management and connectivity**:* 100% of the surveyed institutions have internet connection and a website, 71% have an IT area, and 61% have access to their internal network from every place of the facility.

Infrastructure and availability: 6 out of 10 healthcare providers already conduct some part of the clinical health record in a digitalized format.

ICTs availability for health professionals: The analysis concludes that 98% use a computer at home and 96% have internet access. In addition to this, 50% always use a computer at their working place. Perception of a positive potential impact in the reduction of medical errors, patient adherence to treatment, waiting lines and services waiting lists is less than 60%. However, there is an overall positive perception (70%) on how ICTs impact on the efficiency and service quality improvement, as well as the quality of care as a whole.

Barriers to adoption: The two main barriers identified that can constrain EHR´s adoption were: information security confidentiality (55%), and usability features of EHR (52%).

Discussion

Availability of management and administrative data is higher than clinical information availability. However, clinical information is more commonly used by healthcare professionals. The main results showed that there is a favourable context to advance the use of ICTs in the Uruguayan healthcare system.

Conclusion

Despite measured improvements in ICTs application at health institutions, there are still having challenges and barriers to face the EHR´s adoption at national level. Nevertheless, there would be a promising environment and both a political and technological enabling context for the health sector to successfully implement EHR in Uruguay.

References

[1] OECD Guide to Measuring ICTs in the Health Sector - The Organisation for Economic Co-operation and Development. Retrieved from : http://www.oecd.org/els/health-systems/measuring-icts-in-the-health-sector.htm [Accessed Feb 18th, 2014] .

[2] ICT in health 2013: survey on the use of information and communication technologies in Brazilian health care facilities / coordinator Alexandre F.Barbosa. -- São Paulo : Comitê Gestor da Internet no Brasil, 2014. Retrieved from: http://www.cetic.br/media/docs/publicacoes/2/tic-saude-2013.pdf [Accessed Nov 29th, 2014]

Address for correspondence

cecilia.muxi@agesic.gub.uy

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-922

Syntactic and Semantic Errors in Radiology Reports Associated With Speech Recognition Software

Michael D. Ringler[a], Brian C. Goss[a], Brian J. Bartholmai[a]

[a] *Department of Radiology, Mayo Clinic, Rochester, MN, USA*

Abstract

Speech recognition software (SRS) has many benefits, but also increases the frequency of errors in radiology reports, which could impact patient care. As part of a quality control project, 13 trained medical transcriptionists proofread 213,977 SRS-generated signed reports from 147 different radiologists over a 40 month time interval. Errors were classified as "material" if they were believed to alter interpretation of the report. "Immaterial" errors were subclassified as intrusion/omission or spelling errors. The proportion of errors and error type were compared among individual radiologists, imaging subspecialty, and time periods using χ^2 analysis and multiple logistic regression, as appropriate. 20,759 (9.7%) reports contained errors; 3,992 (1.9%) contained material errors. Among immaterial errors, spelling errors were more common than intrusion/omission errors (P<.001). Error proportion varied significantly among radiologists and between imaging subspecialties (P<.001). Errors were more common in cross-sectional reports (vs. plain radiography) (OR, 3.72), reports reinterpreting results of outside examinations (vs. in-house) (OR, 1.55), and procedural studies (vs. diagnostic) (OR, 1.91) (all P<.001). Dictation microphone upgrade did not affect error rate (P=.06). Error rate decreased over time (P<.001).

Keywords:

PowerScribe, quality control, radiology report, report errors, speech recognition.

Introduction

Speech recognition software (SRS) decreases report turnaround time, but also increases the frequency of semantic and syntactic errors (1). Such errors may negatively impact clinical management, and result in lawsuits (2). We aimed to analyze percentages and types of errors among speech recognition software–generated radiology reports as part of our quality control program.

Materials

We retrieved transcriptionist audits of all speech recognition software–generated radiology reports (PowerScribe; Nuance Communications, Inc.) that were self-edited and signed by 147 different radiologists, between January 3, 2011, and April 16, 2014. 13 trained medical transcriptionists with 1-23 years of experience proofread a random 5% sample of each radiologist's reports each month. Errors were classified as "material" if the transcriptionist could not readily determine the intended meaning in the context of the report. Otherwise, "immaterial" errors were subclassified as intrusion/omission or spelling errors. Reports with multiple errors were only counted once and classified by most egregious error type (material>intrusion/omission>spelling). Punctuation errors were ignored. The proportion of errors and error type were compared among individual radiologists, imaging subspecialty, and time periods using χ^2 analysis and multiple logistic regression, as appropriate.

Results

Of 213,977 reports identified, 20,759 (9.7%) had errors; 3,992 (1.9%) contained material errors. Among 16,767 immaterial errors, spelling errors (10,151, 60.5%) were more common than intrusion/omission errors (*P*<.001). Proportion of errors and fraction of material errors varied significantly among radiologists and imaging subspecialties (*P*<.001). Errors were more common in cross-sectional reports (vs. plain radiography) (OR, 3.72), reports reinterpreting results of outside examinations (vs. in-house) (OR, 1.55), and procedural studies (vs. diagnostic) (OR, 1.91) (all *P*<.001). Dictation microphone upgrade did not affect error rate (*P*=.06). The total error rate also decreased over the 40-month period of this quality control project by 4.3% (*P*<.001).

Conclusion

Speech recognition software–related errors are highly variable among radiologists and imaging subspecialties. One factor may be length of the report. We calculated a relative risk of error in a cross-sectional report of 3.4 (compared to xray),which is consistent with others (3), though we did not specifically track length. Spelling errors are the most common type of immaterial error in self-edited reports, suggesting many radiologists prefer typing over SR. A quality control program with regular feedback may reduce errors over time.

References

[1] Prevedello LM, Ledbetter S, Farkas C, Khorasani R. Implementation of speech recognition in a community-based radiology practice: effect on report turnaround times. J Am Coll Radiol. 2014 Apr;11(4):402-6.

[2] Physician Insurers Association of America. PIAA data sharing reports. Rockville (MD): Physician Insurers Association of America; Jan 1, 1985-Jun 30, 2003.

[3] Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imaging. 2011 Aug;24(4):724-8.

Address for correspondence

Michael D. Ringler, MD, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-923

Diagnostic Imaging Integrated Network: A Teleradiology Pilot in Public Hospitals in Uruguay

Lilian Acosta[a], Karime Ruibal Faral[a]

[a] *Salud.uy, AGESIC, Uruguay*

Abstract

A network of health care centers allows radiologists to share diagnostic images in different areas of Uruguay. This solution is based on an informatics multicenter application for center management, enabling the creation, storage and distribution of images and reports from different imaging modalities according to outsourcing agreements. The solution improves health care territorial equity and reduces asymmetry in resources distribution.

Keywords:

Imaging network; Teleradiology.

Introduction

In 2013 Salud.uy Program designed a pilot project for Teleradiology implementation in Uruguay, developed jointly with an informatics company (MEGA-DATASEC). The Diagnostic Imaging Integrated Network (RIDI in Spanish) was launched in 2014 and aimed at providing a management system for imaging departments and allowing the exchange of imaging studies between healthcare centers.

Methods

The project was conducted in four public hospitals across the country, selected by the Public Health Providers Network (RIEPS in Spanish) [2]. The number of inhabitants, resource complexity and availability of staff in the location were considered. Salud.uy Teleradiology pilot is a system of components and subcomponents of varying degrees of complexity of information hosted on a "cloud". Four components are essential to its goal of achieving a consistent and safe solution for outsourcing images between different centers:

1. Radiological Information System (RIS) as a Policy Enforcement Point -a "door" to the Teleradiology platform- in a Cross-Enterprise Document Sharing (XDS) environment,
2. Administrator Users and Groups for access control,
3. a Picture Archiving and Communication System (PACS) encapsulated in the RIS and
4. a multifunctional proxy appliance or PCS (Picture Communication System) connecting Red.uy network to each local network.

The solution combines a report editor with a web-based viewer for image display and templates of reports according to medical image protocols. The reports are stored in Clinical Document Architecture (CDA) encapsulated in standard Digital Imaging and Communications in Medicine (DICOM) objects [3] to allow future integration with Electronic Medical Record (EMR) and Oncology Electronic Medical Record of Uruguay (HCEO in Spanish) in an IHE XDS environment [4]. In addition, regular monitors were installed in different clinical departments at each hospital, in order to reduce the use of hardcopies. Health professionals and administrative staff were duly trained before RIDI installation.

Results

In Phase 1, RIDI was installed in two centers (the capital and a small provincial city) in September 2014. During Phase 2, two other centers (small provincial cities) are in the process of installation. All centers share the same information management solution (RIS) coupled to a repository (PACS) using a platform for data traffic (Red.uy). RIDI is an open source solution based on interoperability standards and is configured according to the requirements of each hospital. The workflow created by this application impacts the way of working at the imaging department. During the implementation process, resistance to change was detected from the beginning in some of the centers and is currently being solved by providing information, continuous training opportunities and improving usability. Since the project is still in progress, the resulting data are not yet available.

Conclusion

This innovative project allows imaging services in Uruguay to develop collaborative work between centers and at the same time improve each organization. Less patient transfers and shorter waiting times as well as optimization of resources and staff availability across the country enables access to better health services. RIDI also improves doctor-patient relationship by ensuring data safety and confidentiality. Successful development of outsourcing across the country will depend on formal agreements among centers. Furthermore, this project increases and strengthens the sustainability of the National Healthcare System [5]. Finally, once fully achieved, this solution will be made available as a public software.

References

[1] Official paper from Salud.uy Program
[2] Law 18719 of Uruguayan Republic http://www.parlamento.gub.uy/leyes/AccesoTextoLey.asp?Ley=18719&Anchor=
[3] DICOM http://medical.nema.org/standard.html
[4] IHE http://www.ihe.net/Technical_Framework/
[5] Law 18211 of Uruguayan Republic http://www.parlamento.gub.uy/leyes/AccesoTextoLey.asp?Ley=18211&Anchor=

Address for correspondence

lilian.acosta@agesic.gub.uy, karime.ruibal@agesic.gub.uy

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-924

First Step to Big Data Research in Hospital

Yongdon Shin[a], Changmin Choi[a,b], JaeHo Lee[a,c,d], Soo-Yong Shin[a,d]

[a] *Office of Clinical Research Information, Asan Medical Center, Seoul 138-736, Korea.*
[b] *Department of Pulmonary and Critical Care, Asan Medical Center, University of Ulsan College of Medicine, Seoul 138-736, Korea.*
[c] *Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul 138-736, Korea.*
[d] *Department of Biomedical informatics, Asan Medical Center, Seoul 138-736, Korea.*

Abstract

Hospitals have accumulated large amounts of data driven by hospital information system such as EMR, PACS, CPOE, and LIMS. While most data are stored in hospital systems, researchers have still experienced trouble to use clinical data. To overcome this problem and promote "big data" research, clinical research information system is necessary. Here an example of such a system, ABLE (Asan BiomedicaL research Environment), will be introduced.

Keywords:

Clinical Research; Information Retrieval System; Hospital Information System

Introduction

As a lot of information from computer or mobile devices has been produced, "big data" has been highlighted due to its promising economic value. Big data does not only mean collecting a big amount of data, but also investigating meaningful information from the data. To promote big data research in hospitals, researchers should be able to easily search for and extract the necessary clinical data from a Hospital Information System (HIS) to find new information. To utilize hospital clinical data more powerfully in the research sector, we built a research information system called Asan Biomedical research Environment (ABLE).

Methods

ABLE was developed over 6 months (from June, 2013 to December, 2013) by collaborating with Inbrein Co. and Microsoft (MS). The MS .Net framework was used to implement ABLE, and the MS Analytics Platform was chosen as the data warehouse appliance. To build a big data platform, we first carefully chose the research data from HIS. Second, we developed a research data server without building a data mart, since data marts are hard to design and maintain as legacy systems change. Next, legacy data were migrated to the research system by the ETL and Change Data Capture (CDC) processes. Finally, we developed three tools: cohort discovery tool, anonymized chart review tool, and data extraction tool.

Results

The cohort discovery tool is used to determine how many patients in a hospital satisfy the inclusion and exclusion criteria. Anonymized chart review tool supports review of each patient's clinical data by removing personal health identifiers [1]. The data extraction tool supports download of the necessary data in Excel format.

Table 1 is a brief summary of the size of data in ABLE. The number of images seems to be smaller than other clinical data since only CT and MRI images were included in ABLE. Those images are easily de-identified using Digital Imaging and Communications in Medicine (DICOM) protocol. We are developing a new method to de-identify non-DICOM medical images such as sonogram and endoscopy images.

Table 1– The brief summary of data in ABLE

Contents	Size of data
Total registered patients	More than 4M
Total Orders	More than 677M
Total medication	More than 202M
Total laboratory results	More than 740M
Total clinical notes	More than 268M
Total images	More than 22M

Since the launch of ABLE (March, 2014), 239 users accessed ABLE and 4,232 queries were executed per month on average. Before the development of ABLE, researchers requested research data from the office of clinical information. There were about 1,800 data extraction requests per year, and it required 3 - 15 days for researchers to receive the data. However, researchers can search and extract the necessary data on one's own desktop using ABLE. As a result, the total number of data requests was reduced to about 1,390 in 2014.

Conclusion

By implementing ABLE, which follows the research process (i.e., cohort identification, chart review, and data extraction) and protects patients' privacy by removing identifiable information, researchers can easily design research hypotheses, review clinical data while protecting patient's privacy, and extract the necessary data quickly and easily. Therefore, clinical research information systems such as ABLE can be a first step to support big data research in hospitals.

References

[1] Shin SY, Park YR, Shin YD, Choi HJ, Park J, Lyu Y, Lee MS, Choi CM, Kim WS, Lee JH., A De-identification method for bilingual clinical texts of various note types. J Korean Med Sci.: 2015 (1): 7-15.

Address for correspondence

Soo-Yong Shin: Department of Biomedical Informatics, Asan Medical Center, Seoul 138-736, Korea. (Tel) +82-2-3010-2528, (Email) sooyong.shin@amc.seoul.kr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-925

Implementing Georeferencing in the Decision-Making Process of a Health Care Provider in Uruguay

Elizabeth Silva Layes[a], Fernando Morales[b], Julio Leivas[a,b]

[a] *Department of Computing, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay*
[b] *Department of Health Statistics, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay*

Abstract

Proper management of information is not only essential but also critical in any organization's decision making process. Data integration and spacial visualization are the key features of georeferencial systems, which bring support to decision making in epidemiology as well as sanitary planning.This paper shows the development of a georeferential system that interacts with the data provided by the Health Care System.Thereby allowing creation of viable methods which will help to improve patient care, treatments, and interventions through building space-time relationships.

Keywords:

Epidemiology; Georeferencing; Decision support systems.

Introduction

The link between epidemiology and planning of health systems and services is considered of superlative importance. Implementation of such a system becomes essential when meeting the needs and demands of the population through the collection, processing and analysis of data [1]. This will describe the health status of the population and its temporal trends, singling out individuals at risk.

Methods

The application has been developed using a Powerbuilder platform as well as google maps for the first tests of georeferencing.

The user selects the type of analysis to perform, and then sets the values of the variables related to the proposed study. The patients who meet the established criteria will be marked on the map. If required, the system will allow navigation through the patient's EHR, only permitted by security controls.

Results

The developed system allows the visualization of spatial and temporal relationships. The system takes into account different kinds of variables, such as health, socioeconomic, financial and demographic, revealing interrelationships between these variables that help in data analysis.

The system can also support decision making processes in epidemics detection; planification programmes aimed to prevent, carry out surveillance, monitor illnesses and identify potential locations for new health care installations, for example.

Figure 1 shows the results obtained after haven selected female patients with tumors (neoplasms), aged 45 to 64 years, located in the city of Salto, Uruguay for analysis .

This application facilitates the visualization of the location of patients and their associated diagnoses, highlighting the geographic areas of highest incidence.

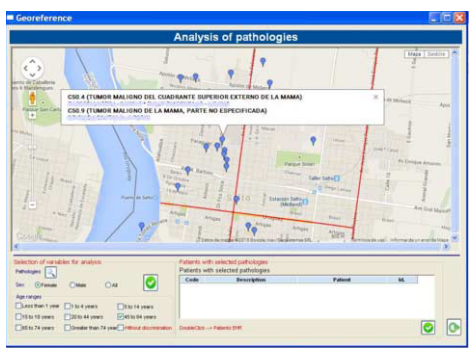


Figure 1 – Female patients with tumors (neoplasms), between 45 and 64 years.

Conclusions

Georeferencing is an important aspect in the study of epidemiology, allowing analysis of geographical and temporal distribution of health events and diseases.

Today it is considered an essential tool, not only in analysis of information for epidemiological decision-making, but also to properly manage various resources in a health care organization. In this sense, the developed application allows analysis of different scenarios through inclusion of multiple variables.

References

[1] Bhatt BM, Joshi JP: GIS in epidemiology: Applications and services. Natl J Community Med. 2012;3(2):259-263.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-926

A Development of Automatic Audit System for Written Informed Consent using Machine Learning

Hitomi Yamada[ab], Tadamasa Takemura[b] Takahiro Asai[b], Kazuya Okamoto[c], Tomohiro Kuroda[c], Shigeki Kuwata[a]

[a] *National Cerebral and Cardiovascular Center, Suita, Japan*
[b] *Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan*
[c] *Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan*

Abstract

In Japan, most of all the university and advanced hospitals have implemented both electronic order entry systems and electronic charting. In addition, all medical records are subjected to inspector audit for quality assurance. The record of informed consent (IC) is very important as this provides evidence of consent from the patient or patient's family and health care provider. Therefore, we developed an automatic audit system for a hospital information system (HIS) that is able to evaluate IC automatically using machine learning.

Keywords:
Medical Chart, Automatic Audit System, Machine Learning.

Introduction

Medical records provide evidence to protect both patient and health care worker's rights [1]. Informed consent (IC) is important as it provides evidence of the health care worker's obtaining patient consent, which is critical to health care. If health care workers cannot agree about their care in practice, they faced a heightened risk. The health care workers have to record their agreement in just proportion and hospital administrators have to audit said records in order to protect their mutual interests. It is important to audit patient medical record records to verify written IC has been obtained. However it is very difficult to audit all electronic medical records by humans alone. Therefore, we discussed the possibility of an automatic audit system in order to quality check the medical records for IC using a machine learning technique [2].

Materials and Methods

First, we extracted 298 electronic IC charts from HIS at the National Cerebral and Cardiovascular Center (NCVC) in Japan. At this facility, all medical documentation is elecronically recorded into the HIS by health care workers. A health information manager evaluated the IC documents in five levels from an expert point of view. Next, we considered what level was adequate for IC content, and we determined upper level 3 was adequate quality and lower level 2 was not. We did morphological analysis in Japanese using a medical dictionary to detect specific medical terms in the IC charts. We also used medical words and word numbers as training examples for machine learning. Concretely, we used support vector machines (SVM's) for supervised learning models to be able to mark IC documents as belonging to one of the two categories. At this time, the system determined whether the IC document was a lower level 2 or not.

Results

We evaluated this system by cross-validation. Specifically, we evaluated this system using leave-one-out validation (LOOCV). Consequently, this system could evaluate the IC documents 89.4% (261/292) correctly. The false negative rate was 29% (16/56) and false positive rate was 6.4% (15/236) in determing the lower level 2 IC charts.

Conclusions

A quality review of IC documents is very important in order to protect health care worker and patient rights. The results of this research suggest a possible benefical effect of our automated method utilizing SVM and machine learning. In addtion, we have not adjusted parameters of the SVM yet so it is still possible that this system will learn more data. Additional improvements to this system will need to be made for clincal practice.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26460869.

References

[1] A.R. Bakker, "The need to know the history of the use of digital patient data, in particular the EHR" *Int.J.Med.Info*, vol. 76(5), 2007, pp. 438–441.

[2] T.Takemura, K.Okamoto, et al. "Context-Based Loose Information Structure for Medical Free Text Document" HCII2007, 2007, pp. 544–548.

Address for correspondence

Tadamasa Takemura, Graduate School of Applied Informatics, University of Hyogo, Minatojima-Minamimachi 7-1-28, Chuo-ku, Kobe, Japan.

Email: takemura@ai.u-hyogo.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-927

Service Quality: A Main Determinant Factor for Health Information System Success in Low-resource Settings

Binyam Tilahun, Fleur Fritz

Institute of Medical Informatics, University of Münster, Germany

Abstract

With the increasing implementation of different health information systems in developing countries, there is a growing need to measure the main determinants of their success. The results of this evaluation study on the determinants of HIS success in five low resource setting hospitals show that service quality is the main determinant factor for information system success in those kind of settings.

Keywords:

Service quality, Low-resource settings, IS success, Evaluation

Introduction

Information system (IS) success measurement is a topic of substantial research and debate. A number of researchers have derived various models to explain what makes IS successful and how to measure it [1, 2].

Among them, the Delone and Maclean Model (D&M) is regarded as the most used and appropriate way to measure IS success by many researchers [1]. The basic dimensions in this model are: system quality, information quality, service quality, system use, user satisfaction and net benefit. The model assumes that those constructs are interrelated and interdependent. Among those factors, user satisfaction is regarded as a primary determinant to measure IS success [3].

With the increasing implementation of health IS in developing countries, it is necessary to identify the main determinant factors as an input for policy making and priority setting. Given that in those settings there is usually limited infrastructure available, it is necessary to assess the quality of internal and external service support for proper use of the system. In this paper our objective is to assess the effect of service quality on user satisfation and hence on IS success.

Methods

Service quality from D&M is operationalized as " the quality of the infrastructure and IT desk support in the hospitals". To measure this, we developed a questionnaire based on the validated five dimensions of SERVQUAL [4]. Additionally, we added setting specific questions like power interruption.

The data collection was part of a hospital IS evaluation study among health profesionals in five public hospitals in Ethiopia. The data collection was conducted from January to February, 2014 .

Results

Out of the 422 participants in this study, 406 (96%) completed the questionnaire. The mean age of the participants was 34 years (±8.5). 217 (53.2%) were males and the majority of the participants were nurses 176 (43.3%), followed by physicians 83 (20.4%) and HMIS staff 74 (18.2%). 190 (61%) of the health professionals reported overall dissatisfaction with the EMR (Median=4, IQR=1) on a five-level Likert scale. Physicians were more dissatisfied (Median=5, IQR=1) as compared to nurses (Median=4, IQR=1) and HMIS staffs (Median=2, IQR=1).

The respondents indicated disagreement with service quality of the implemented system with an overall median score in the range of "Disagree" (Median=4.5 IQR=1.5). The logistic regression shows a strong correlation between service quality and satisfaction - those profesionasl who belive the service quality is good are eight times (AOR=8.23 95% CI (3.23-17.01)) more likely to use the system than the others. The association with system quality was two times (AOR=2.2 95% CI (1.34-3.09)) and with information quality was also aproximately two times (AOR=1.94 95% CI (1.12-3.23)).

Among the factors that affect service quality, 222 (72%) think that the IT support staff do not understand their needs, 191 (62%) believe that the initial training was not adequate and 227 (73%) were also unhappy with the frequent power interruption in the study hospitals.

Conclusion

The results of this study show that service quality has a strong correlation with user satisfaction, hence to IS success in low-resource settings. Therefore, system implementation efforts in those settings should focus on improving the service quality such as power infrastructure, IT support and trainings. Further study on the interrelationship between D&M constructs and HIS success in those settings is in progress.

References

[1] Maclean W, and DeLone H. The DeLone and McLean Model of Information Systems Success : A Ten-Year Update. *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003.

[2] Seddon P. A respecification and extension of the DeLone and McLean model of IS success. *Inf. Syst. Res.*, 1997.

[3] Eijden M, Ange HJT, Roost JT, and Asman AH. Determinants of Success of Inpatient Clinical Information Systems. vol. 10, no. 3, pp. 235–243, 2003.

[4] Watson R, Pitt L, and Kavan C. Measuring information systems service quality *MIS Q.*, 1998.

Adress for Correspondance

Binyam.Tilahun@uni-muenster.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-928

Taming the Data Quality Dragon – A Theory and Method for Data Quality by Design

Jens H. Weber[a,b], Morgan Price[a,b], Iryna Davies[b]

[a] *Department of Computer Science, University of Victoria, Victoria, BC, Canada*
[b] *Department of Family Practice, University of British Columbia, Vancouver, BC, Canada*

Abstract

A lack of data quality (DQ) is often a significant inhibitor impeding the realization of cost and quality benefits expected from Clinical Information Systems (CIS). Attaining and sustaining DQ in CIS has been a multi-faceted and elusive goal. The current literature on DQ in health informatics mainly consists of empirical studies and practitioners' reports, but often lack a holistic approach to addressing DQ 'by design'. This paper seeks to present a general framework for clinical DQ, which blends foundational engineering theories with concepts and methods from health informatics. We define an architectural viewpoint for designing and reasoning about DQ. We introduce the notion of DQ Probes for monitoring and assuring DQ during system operation. The concepts presented have been validated in a real-world case study.

Keywords:
Information Management [L01.399]; Engineering [J01.293].

Introduction

The issue of clinical DQ has many facets. There is not yet a single commonly accepted standard definition of this concept; there however has been growing consensus on a number of aspects that have to be taken into account when discussing DQ in a health informatics context. Weiskopf and Weng have mapped DQ issues reported in the CIS literature to five quality attributes (*completeness*, *correctness*, *currency*, *plausibility* and *concordance*) [1]. The published literature on these DQ attributes is primarily of empirical nature or focuses on a particular practical aspect of defining, measuring or improving DQ for a particular purpose [2]. The lack of a fundamental theory and method for engineering CIS for DQ presents an impediment to the design and evolution of better health systems. The purpose of this paper is to define such a theory and method, referred to as *DQ by Design* (DQbD).

Methods

The DQbD method is based on the engineering paradigm of *functional decomposition*, in which complex system behaviour is broken down into more elementary discrete functions. We incorporated the idea of *Design by Contract (DbC)* from component-based software engineering and extended it to enable assertions over DQ concerns. We introduced and defined the concept of *DQ Probes* as a way to make DQ related assertions in pre- and post-conditions associated with CIS data processing functions. Our theory and method have been validated in several CIS data projects, including our project on building a third generation primary care CIS research network [3].

Results

We have formalized a taxonomy of five types of DQ probes:

Table 1 – Taxonomy of CIS Data Quality Concerns

	DQ Type		Clinical Example
0	Intrinsic	Conformance	Is diabetes on the problem list coded or uncoded?
1	Intrinsic-meta	Provenance	Recency of last A1c test in a diabetic
2	Extrinsic-internal	Internal concordance	Do patients on insulin have Diabetes on their problem list?
3	Extrinsic-external	External concordance, currency	Are the same allergies present in the hospital and primary care information systems? Which were updated most recently?
4	Statistical	Plausibility	Is the expected prevalence of diabetes in a practice (as determined in their EMR) for males 55-70 similar to published regional, jurisdictional, or national averages?

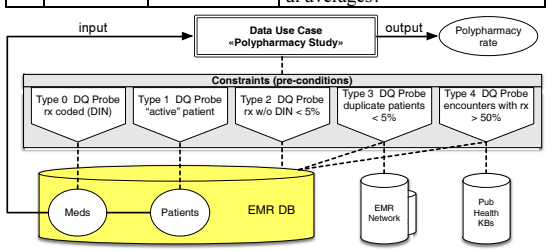


Figure 1 – DQ Probes in pre-conditions of Data Use Cases

We developed an architectural view point for designing CIS system use cases with explicit deployment of DQ probes. Our theory has been validated in a real-world project [3].

References

[1] Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J. Am. Med. Inform. Assoc., 2013; 20(1): 144-151.

[2] Mettler T, Rohner P, Baacke L. Improving data quality of health information systems: a holistic design-oriented approach. Proceedings of the 16th European Conference on Information Systems; Galway, Ireland. 2008: pp. 1883-1893.

[3] Price M, Weber JH, McCallum G. SCOOP–The social collaboratory for outcome oriented primary care. *IEEE ICHI* 2014: pp. 210–215.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-929

A Medical Image Backup Architecture Based on a NoSQL Database and Cloud Computing Services

Luan Henrique Santos Simões de Almeida[a], Marcelo Costa Oliveira[a]

[a] *Laboratório de Telemedicina e Informática Médica, Instituto de Computação, Universidade Federal de Alagoas, Maceió, AL, BR*

Abstract

The use of digital systems for storing medical images generates a huge volume of data. Digital images are commonly stored and managed on a Picture Archiving and Communication System (PACS), under the DICOM standard. However, PACS is limited because it is strongly dependent on the server's physical space. Alternatively, Cloud Computing arises as an extensive, low cost, and reconfigurable resource. However, medical images contain patient information that can not be made available in a public cloud. Therefore, a mechanism to anonymize these images is needed. This poster presents a solution for this issue by taking digital images from PACS, converting the information contained in each image file to a NoSQL database, and using cloud computing to store digital images.

Keywords:
Cloud Computing; NoSQL Database; PACS.

Introduction

The Picture Archiving and Communication System (PACS) is an approach widely used to store digital images. In 2010, Teng et al. [1] stated that in 2014 the United States produced approximately 100 Petabytes of medical images. Nevertheless, PACS is limited in current scenarios, with high investment in data centers and limited physical storage. Therefore, adopting a scalable environment that can persist data as long as needed and withstand varying loads is essential.

Cloud Computing is able to provide an elastic and appropriate environment for a set of applications besides providing hardware and software resources as a service, providing a low-cost storage with high reliability and security [2]. Thus, cloud computing becomes the ideal setting for several health applications, providing a great resource for storage and digital image processing. However, DICOM images contain private patient information; therefore, a mechanism to anonymize these images is necessary. The objective of this poster is to build an automated backup environment of anonymized medical images derived from PACS into a public cloud, powered by a NoSQL database.

Materials and Methods

The experiments were developed using the PACS solution from dcm4che project, the Amazon S3 public cloud and 1761 computed tomography (CT) images in the DICOM standard. The patient information contained in the DICOM file was stored in a MongoDB NoSQL database since these DICOM files are formed by tags that are similar to the key-value documents in MongoDB. As shown in Figure 1, the images are generated by acquisition modality equipments, sent to the PACS, read by the backup manager, and uploaded to the cloud after removing personal information.

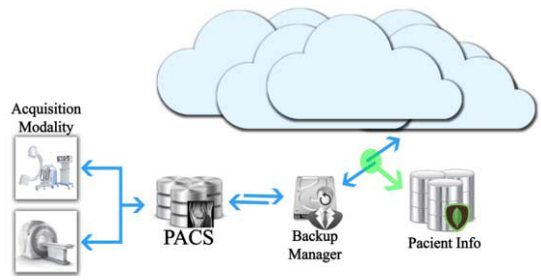


Figure 1 - Architecture overview

Results and Conclusion

The environment was built to be a fail-safe system by using MongoDB documents as log reports. Through these log reports, the environment will be able to recognize in which task the backup process stopped and continue from that point. The reverse way is also possible and is in development. In other words, the system will be able to identify the patient ID from the patient name, recover the file from the cloud, then place the patient information inside this file, and identify them. The query for patient information was optimized due to the use of MongoDB documents instead of regular SQL databases since it is possible to parallelize queries and these documents consisting of key-value pairs. Using a NoSQL database allows integration between other health bases within the big data concept in a near future.

References

[1] Teng CC, Mitchell J, Walker C, et al. A medical image archive solution in the cloud. International Conference on Software Engineering and Service Science (ICSESS), IEEE, 2010; 431-434.

[2] Zeng W, Zhao Y, Ou K, Song W. Research on cloud storage architecture and key technologies. Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, 2009. p. 1044–1048, New York: ACM.

Address for correspondence
luanhssa@ic.ufal.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-930

Digital Imaging and Electronic Data Capture in Multi-Center Clinical Trials

Thomas M. Deserno[a], Verena Deserno[b], Daniel Haak[a], Klaus Kabino[a]

[a]*Department of Medical Informatics, Uniklinik RWTH Aachen, Germany*
[b]*Clinical Trial Center Aachen (CTC-A), Uniklinik RWTH Aachen, Germany*

Abstract

While medical image data is managed in picture archiving and communication systems (PACS) via the digital imaging and communications in medicine (DICOM) protocol, electronic data capture systems (EDCS) in clinical trials lack PACS interfacing. This complicates the trial workflow and increases errors, time, and costs. In this work, four system architectures of image integration for multi-center trials are analyzed with respect to data, function, visual, and context integration levels. We propose an open source-based architecture composed of OpenClinica, DCM4CHE, and Weasis for EDCS, PACS, and Viewer, respectively.

Keywords: Clinical Trials, Clinical Research, Imaging Biomarkers, Image-based Surrogates, Image Management System Integration, Workflow Integration

Introduction

Providing surrogate endpoints in clinical trials, medical imaging has become increasingly important in personalized medical research [1]. Electronic data capture systems (EDCS) are used to record research data while picture archiving and communication systems (PACS) manage subject's imaging data. Despite the digital imaging and communications in medicine (DICOM) protocol, EDCS and PACS are currently not interconnected. Particularly in multi-center trials, manual data interchange yields errors, delays, and additional costs.

Materials and Methods

The clincial PACS is separated from the research PACS, where all DICOM data is de-identified but linked to the subject matrix and electronic case report form (eCRF) in the EDCS. For storage, the research nurse might operate in either system as the leading component. For retrieval, DICOM objects might be viewed via stand-alone DICOM viewer or integratively via Web-based browsers (Figure 1).

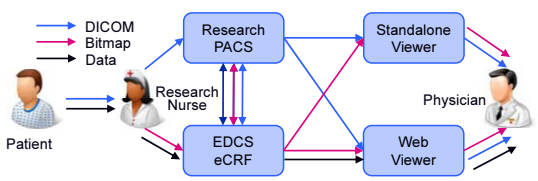


Figure 1

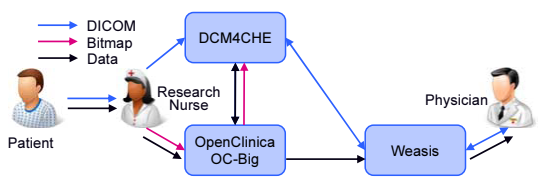
Using EDCS as primary system, there are four system architectures:

1. Image data are stored via EDCS as binary large object (BLOB). Retrieval may be supported via a Web-based DICOM viewer.
2. DICOM data are transferred via EDCS to the research PACS for storage, de-identification, and retrieval.
3. DICOM data are directly sent to the research PACS, and identifiers are handed back to EDCS' subject matrix.
4. Results from manual or automatical image analysis are stored in the PACS (e.g., DICOM Structured Reporting).

Results

For Level 1, EDCS and viewer have to support BLOB and DICOM data, respectively. Level 2 requires DICOM functionality in the EDCS, and both PACS and viewer have to support web access to DICOM persistent objects (WADO). To accept DICOM objects in the PACS directly, appropriate links to EDCS are required. For Level 4, the viewer component must support advanced DICOM services yielding full data, function, visual, and context integration.

We suggest OpenClinica, DCM4CHE, and Weasis as open source components for EDCS, PACS, and Web viewer.



Discussion & Conclusions

The successive levels of EDCS/PACS integration provide increasing functionality in mutli-center trials. At all levels, the EDCS is considered as primary system best supporting the research nurses' workflow. This is in contrast to van Herk, who tightly integrates the medical PACS as primary system, transferring the DICOM identifyiers into the eCRF [2].

References

[1] Mandrekar SJ, Sargent DJ. Drug designs fulfilling the requirements of clinical trials aiming at personalizing medicine. Chin Clin Oncol. 2014 Jun 1;3(2):14

[2] van Herk M. Integration of a clinical trial database with a PACS. J Phys 2014;489(1):1-6.

Address for correspondence

Thomas Deserno, deserno@ieee.org, +49 241 80 88793

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-931

An Exponential Increase in Regional Health Information Exchange With Collaborative Policies and Technologies

N. Lance Downing [a, b], Steven Lane [c], Mathew Eisenberg [b], Christopher Sharp [b], Jonathan Palma [a], Northern California HIE Working Group, Christopher Longhurst [a]

[a] Stanford Children's Health, [b] Stanford Health Care, [c] Sutter Health, Palo Alto Medical Foundation

Abstract

In the United States, the ability to securely exchange health information between organizations has been limited by technical interoperability, patient identity matching, and variable institutional policies. Here, we examine the regional experience in a national health information exchange network by examining clinical data sharing between eleven Northern California organizations using the same health information exchange (HIE) platform between 2013-2014. We identify key policies and technologies that have led to a dramatic increase in health information exchange.

Keywords:

Electronic Health Records; Health Information Exchange.

Introduction

As part of the Meaningful Use program, the U.S. government legislated incentives and penalties to encourage HIE, citing reasons of patient safety, potential cost savings, and patient engagement. Thus far, widespread adoption and benefits have been limited, though there have been promising results in reducing emergency department usage of redundant diagnostic testing and unnecessary hospital admissions. Here, we describe policy and technology changes associated with a dramatic increase in regional secure health information exchange across a national HIE network.

Methods

11 of the 12 health systems in Northern California that use the same electronic health record (EHR) and integrated HIE vendor (Epic Systems, Verona, WI) provided their regional HIE data. We examined the HIE activity between November, 2013– November, 2014 as measured by the number of clinical summaries received by each institution from the other 10 during this period. Additionally, we identified key policy and technology changes during this timeframe that may have contributed to the extraordinary increase in secure health information exchange.

Results

A steady increase in HIE volume is seen throughout this period with a marked increase in certain institutions at specific time points (Figure 1).

Discussion

We observed a dramatic increase in secure HIE during this period that coincides with changes in institutional policies and HIE functionality. Patient authorization requirement policies differ between these institutions, ranging from none, only if the patient has an encounter in a confidential department, or always at the point of care. During the 2014 period, most of the organizations adopted open HIE policies; waiving patient authorization requirement unless specifically restricted by the patient or clinical context.

Additionally, the institutional policies to automatically query regional organizations for patient matches varies, with options for permitting both sending and receiving queries. Changes to both the authorization requirement policy and auto-query functionality coincided with a dramatic increase in HIE volume.

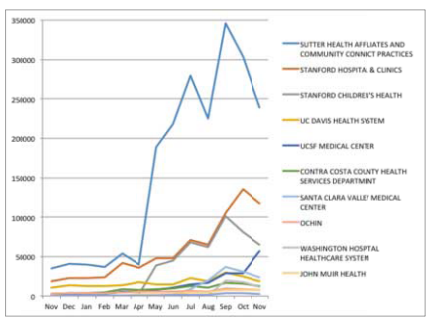


Figure 1 – Clinical Summaries Received by Organization

Conclusion

A striking increase in secure HIE is seen in this regional use of a single vendor HIE network due to changes in patient authorization policies and the use of new automated HIE functionality.

References

[1] Rudin, R. S., Motala, A., Goldzweig, C. L., & Shekelle, P. G. (2014). Usage and Effect of Health Information Exchange: A Systematic Review. *Annals of Internal Medicine*, *161*(11), 803–811.

[2] Lammers, E., Adler-Milstein, J., & Kocher, K. (2014). Does Health Information Exchange Reduce Redundant Imaging? Evidence From Emergency Departments. *Medical Care*, *52*(3), 227–234.

Address for correspondence

N. Lance Downing, M.D., Stanford Children's Health. 4100 Bohannon Dr., Menlo Park, CA 940

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-932

Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the Open Medical Record System

Suranga N. Kasthurirathne[a], Burke Mamlin[b,c], Grahame Grieve[d], Paul Biondich[b,c]

[a] School of Informatics and Computing, Indiana University – Purdue University, Indianapolis, IN, USA, [b] Indiana University School of Medicine, Indianapolis, IN, USA, [c] Regenstrief Institute Inc., Indianapolis, IN, USA, [d] Health Intersections Pty. Ltd. , AU

Abstract

Interoperability is essential to address limitations caused by the ad hoc implementation of clinical information systems and the distributed nature of modern medical care. The HL7 V2 and V3 standards have played a significant role in ensuring interoperability for healthcare. FHIR is a next generation standard created to address fundamental limitations in HL7 V2 and V3. FHIR is particularly relevant to OpenMRS, an Open Source Medical Record System widely used across emerging economies. FHIR has the potential to allow OpenMRS to move away from a bespoke, application specific API to a standards based API. We describe efforts to design and implement a FHIR based API for the OpenMRS platform. Lessons learned from this effort were used to define long term plans to transition from the legacy OpenMRS API to a FHIR based API that greatly reduces the learning curve for developers and helps enhance adhernce to standards.

Keywords:

HL7; FHIR; Interoperability; Standards; Electronic Medical Records.

Introduction

Healthcare quality is enabled by the accessibility and effective use of clinical data. The specialized nature of medicine and the ad hoc implementation of clinical information systems have led to the fragmentation of health information and medical care [1]. These limitations spurred efforts to effectively exchange patient data between various stakeholders across the enterprise, and led to the development of HL7 V2 and V3 standards. The FHIR standard is a next generation standard framework introduced in response to limitations in HL7 V2 and V3 [2]. FHIR has much to offer OpenMRS, a mature Open Source EMR system that is widely used across resource limited settings [3]. This paper describes efforts to design and develop FHIR support for OpenMRS. We demonstrate the feasibility and value of adopting FHIR for OpenMRS and develop an architectural plan to transition from the legacy OpenMRS API to a FHIR specific API.

Methods

We leveraged OpenMRS's modular architecture to develop an add-on module to provide a FHIR based API for OpenMRS. The FHIR module integrates the HAPI FHIR API for modeling and populating FHIR resources. The module exposes a FHIR API that could be accessed internally via API calls, or externally using FHIR specific user requests. The FHIR API receives user requests and serves them by calling the domain specific legacy OpenMRS API.

Results

We developed a production worthy FHIR module that could be used to fulfill basic user requirements for exchanging clinical data. The initial version of the module supports the import / export of OpenMRS clinical data in the form of FHIR standards for person, patient, location, observation, allergy intolerance, condition and encounter resources. The module is able to import / export FHIR resources in either xml or json. The module can also manage changing versions of OpenMRS domain objects via a strategy pattern that populates FHIR resources based on the underlying version of the OpenMRS platform. Each FHIR resource complies with the latest FHIR specification, which stood at Draft Specification for Trial Use (DSTU) 2 at time of submission.

Discussion

We perceive that the FHIR API will be widely adopted as it is based on a hugely popular standard, and removes the need for system developers to learn the OpenMRS domain model. However, we anticipate that implementer uptake will be a slow process that will happen gradually overtime. The FHIR module will initially be installed in parallel with the existing domain specific API. However, we propose a way forward to gradually retire the legacy OpenMRS API in favor of the standardized FHIR API. These efforts represent proper separation of concerns between EMR developers and champions of interoperability, and enables better interoperability across a wider range of healthcare IT infrastructure.

References

[1] Stange KC. The problem of fragmentation and the need for integrative solutions. The Annals of Family Medicine. 2009;7(2):100-3.

[2] Bender D, Sartipi K, editors. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on; 2013: IEEE.

[3] Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, et al. The OpenMRS implementers network. International journal of medical informatics. 2009;78(11):711-20.

Address for correspondence

S.N. Kasthurirathne, Indiana University School of Informatics and Computing, 535 W. Michigan Street, IT 475, Indianapolis, IN 46202, USA, (317)-278-4636, snkasthu@iupui.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-933

Development of Clinical Database System Specialized for Heavy Particle Therapy

Masami Mukai[a], Yutaka Ando[b], Yuki Yokooka[a], Yasuo Okuda[a], Masayoshi Seki[c], Masahiro Kimura[d], Hiroshi Tsuji[b], Tadashi Kamada[b]

[a] Medical Informatics Section, National Institute of Radiological Sciences (NIRS), Chiba, Japan
[b] Research Center for Charged Particle Therapy, NIRS, Chiba, Japan
[c] Global-for Co., Tokyo, Japan, [d] Fujitsu Systems East Limited, Saitama, Japan

Abstract

We have developed a data archiving system for study of charged particle therapy. We required a data-relation mechanism between electronic medical record system (EMR) and database system, because it needs to ensure the information consistency. This paper presents the investigation results of these techniques. The standards in the medical informatics field that we focus on are Integrating the Healthcare Enterprise (IHE) and 2) Health Level-7 (HL7) to archive the data. As a main cooperation function, we adapt 2 integration profiles of IHE as follows, 1) Patient Administration Management (PAM) Profile of IHE-ITI domain for patient demographic information reconciliation, 2) Enterprise Schedule Integration(ESI) profile of IHE-Radiation Oncology domain for order management between EMR and treatment management system(TMS). We also use HL7 Ver2.5 messages for exchanging the follow-up data and result of laboratory test. In the future, by implementation of this system cooperation, we will be able to ensure interoperability in the event of the EMR update.

Keywords:

Radiotherapy, Database, Strandards, IHE, HL7.

Introduction/Purpose

Our hospital has a mission of clinical research for radiotherapy. Charged particle therapy (carbon ion) was started in 1994, and over 9,500 cases have been treated by November, 2014. To accomplish this mission, we managed multi-system such as electronic medical record systems (EMR) and charged particle therapy treatment management system (TMS).

In 2000, we started to operate the Advanced Medical Information Database System (AMIDAS) for archiving the radiotherapy information. With the starting of EMR, we allocated a role to information systems as follows, EMR: input data related radiotherapy, AMIDAS: make report and summary of radiotherapy. So the AMIDAS is required to construct a mechanism to collect the data which is input by end-user on EMR.

Methods

The data targeted for the cooperation are following: (1) patient demographic information, (2) tumor related information, (3) radiation plan information, (4) follow-up information (tumor effect, advance reaction, mortality, etc.), (5) laboratory results, (6) treatment delivery information. We divided the implementation process into two stages and examined it as two steps: (1) investigated the availability of IHE [1]. (2) investigated the use of HL7 messages.

Results

This cooperation function was realized by two IHE integration profiles as follows, (1) Patient demographics and visit information: PAM Integration Profile, (2) Radiotherapy order and delivery information: ESI Integration Profile. For communication of treatment follow-up information and laboratory test we defined context and used HL7 messages.

Discussion

We show the comparison results using standard with original sysytem-interface in Table 1.

Table 1– The Comparison of Standard with original messages

Comparisonpoint	Standard-IHE	Standard-HL7	Orignal interface
Meeting number of times	little	few	much
The use of the library	possible	possible	impossible
Time to make specifications	short	middle	long

Conclusion

In comparison with original message system interface, it may be said that the system which was developed using a standardization technology has interoperability. From the standpoint of system-operation by using standards, when we will renew the EMR, AMIDAS can receive the data from EMR without software modification.

References

[1] IHE(Integrating the Healthcare Enterprise) http://www.ihe.net/Technical_Frameworks/

Address for correspondence

MUKAI Masami E-mail:m_mukai@nirs.go.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-934

Interoperable Archetypes With a Three Folded Terminology Governance

Rune Pederson[a,b1], Gunnar Ellingsen [a,b]

[a] *University Hospital of North Norway, Tromsø, Norway*
[b] *Telemedicine and eHealth research group, Arctic University of Norway, Norway*

Abstract

The use of openEHR archetypes increases the interoperability of clinical terminology, and in doing so improves upon the availability of clinical terminology for both primary and secondary purposes. Where clinical terminology is employed in the EPR system, research reports conflictinga results for the use of structuring and standardization as measurements of success. In order to elucidate this concept, this paper focuses on the effort to establish a national repository for openEHR based archetypes in Norway where clinical terminology could be included with benefit for interoperability three folded.

Introduction

Initiatives using an openEHR architecture have been established for two primary purposes: (a) to build a national repository of common semantic data elements for collaborative EPR systems, and (b) to serve a large EPR vendor in building their system portfolio around the openEHR technology. The Clinical Knowledge Manager for archetypes in Norway is scheduled to contain between 1000 and 2000 archetypes, archiving information about how new archetypes are translated, modelled, and shared. This will give a common understanding of the clinical content of EPR systems over regional, national and global boundaries coordinated by the National Administration Office of Archetypes (NRUA). The increased focus on process-oriented systems across different health care organizations also presupposes standardization in the form of shared terminologies like NIC/NANDA, ICD, SNOMED-CT, and ICNP. This will generate an automatic and reliable use of terminology for sharing semantic interoperable information. By operating openEHR archetypes and terminology integration, it is possible to structure EPR content in a multilevel modelling approach that includes templates, archetypes, and a reference model. Within this structure, reuse of EPR data for research purposes and sharing information between heterogeneous practices become possible.

Results

Throughout the last three years, the use of OpenEHR archetypes has grown to gain a national anchorage in Norwegian healthcare. From the outset, a national collaborating group is coordinating with the vendor to build a national repository of OpenEHR archetypes. At the time it seems obvious that clinical terminologies are the connecting factor between the work with archetypes and the Norwegian Directorate of Health's focus on terminology work. With this as a backbone we observe that when clinical variables become structured and the semantic interoperability is governed by the possibilities within the OpenEHR platform, the sharing of information over local, national and global boundaries, for primary and secondary use of information, gain new possibilities.

Method

Qualitative methods (interpretive) have been applied, grounded in the first author's participation and contribution in the work accomplished. observations and description of ongoing work has been followed by interviews with members of the regional and national initiatives, which so far includes 9 interviews arround the archetype governance, several interviews and observations on the use of clinical terminology, and conversations with end users of the Clinical Knowledge Manager while guiding them to become users.

Discussion

Health authorities worldwide have faced a growing pressure to accomplish a smooth information flow among multi-organizational health systems. The motivation for this stems from the need for a streamlined work processes across organizational boundaries, and from the need to ensure quality of treatment and care of patients and secondary use purposes. To avoid mediating problems to reach primary and secondary target it becomes essential that standardized terminologies for different domains can be integrated to structured data elements in the EPR system. Although terminology standards are used on a daily basis in health care work, we know little about the processes of how these terminologies come into being and about how they are co-constructed with daily work.

Conclusion

Equal for both archetypes and terminology is that information used for any given purpose on a local or national level can be reused with another purpose both locally, nationally, and globally (the three levels). For archetypes that are a) approved in the national governance processes, b) increasing in number, or c) comprise clinical support repositories, there will be opportunity for terminology binding. An increasing number of archetypes, more than 1000, will in the end be accessible in an open repository. Further, each archetype that is translated or modelled will be compared or reviewed with the purpose of being added to the global repository.

[1] Corresponding Author.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-935

Standardization of Information about Birth in the Obstetric Discharge Summary

Zilma S Nogueira Reis[a], Juliano S Gaspar[a], Isaias J R Oliveira[a], Andreia C de Souza[b], Thais A Maia[c]

[a] *Faculty of Medicine, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil*
[b] *Postgraduate Studies in Technology in Health, Pontificia Universidade Católica do Paraná, Maringá, Brazil*
[c] *Postgraduate Studies in Information Systems and Knowledge Management, Universidade FUMEC, Belo Horizonte, Brazil*

Abstract

Clinical information about the birth composes an important set of data to the documentation about the care provided during childbirth. Formalized in the document Obstetric Impatient Discharge Summary (OIDS), such information are essential for continuity of mother and child attention, in the health care network. The main paper's objective is to propose an Information Model for this document based on ISO Standard 13606 for interoperability between health information systems in Minas Gerais, Brazil.

Keywords:
Patient Discharge; Continuity of Patient Care; Health Information Systems.

Introduction

The OIDS is a strategic document able to collaborate in the construction of integration and interdependence between the hospital and primary care, using computerized communication channels. This study is a pilot that intends to prepare electronic systems of public information, already operational, for the challenges of transmission of clinical data.

Methods

This is a qualitative study with methodological exploratory purposes, which seeks to systematize and standardize the reference document model, aiming at interoperability between electronic information systems in Minas Gerais, Brazil. The set of clinical data about childbirth was proposed with the formalities and informational structure necessary to preserve the semantics of the data.

Results

Starting from a set of clinical data previously proposed by domain experts[1] and the archetypes published by Health Department of the State of Minas Gerais [2], the clinical data set was reviewed. In order to address many possibilities of obstetric hospitalization, the information model was divided into two sections: one about women's health, and other about the newborn. Based on archetypes of ISO Standard 13606, the knowledge model was formalized using the tool LinkEHR Platform[3]. All the documents related to the information model are available on the project website: http://site.medicina.ufmg.br/cins/apoio-a-pesquisa-22/projetos/.

Discussion

The harmonization of essential information defined by clinical experts with the existing on the web page that contains the archetypes published by the Health Department of Minas Gerais resulted in the reuse of two existing archetypes and the creation of ten new ones. However, no adaptation was necessary for terminology. SISMater® is a system used to collect obstetric and neonatal clinical data in a maternity of a public hospital and the data from it will be used to complete the proposed information model to send information to the basic health units systems for continuity of patient care.

Conclusion

It is considered that the study achieved its objective, since the information model for the OIDS based on the ISO Standard 13606 was created. It is expected that the document ensure semantic interoperability to facilitate communication on health care network for the continuity of patient care and its clinical safety. As the future study, it is suggested to apply the information model in different maternities.

Acknowledgments

Supported by FAPEMIG – PPSUS Redes, APQ. Endorsed by the Workgroup 1 from ABNT – Brazil.

References

[1] Ferreira AAT, Reis ZSN, Frade S, Gaspar JS, Correia RJC, Hadad SC, et al. Proposição de um sumário de alta obstétrico visando à troca de informações, em padrão OpenEHR, para continuidade do cuidado materno-infantil. Medicina (Ribeirão Preto). 2014:47(Supl 1):59-66.

[2] Minas Gerais. SES. Portal Público B-RES. 2012. Available from: sres.saude.mg.gov.br/arquetipo/listar.

[3] Maldonado JA, Moner D, Boscá D, Fernández-Breis JT, Angulo C, Robles M. LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. International journal of medical informatics 2009:78(8):559-70.

Address for correspondence

Av. Prof. Alfredo Balena, 190. Sala 606. CEP 30.130.100 Belo Horizonte. Minas Gerais. Brazil. E-mail: zilma@medicina.ufmg.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-936

Hospitalization Discharge Summary: Standardization of Information Model

Andréia Cristina de Souza[ab], Claudia Moro[a], Zilma S N Reis[b]

[a]*Graduate Studies in Technology in Health, Pontifical Catholic Univeristy of Paraná (PUCPR), Curitiba, Brazil*
[b] *Faculty of Medicine, Federal University Minas Gerais (UFMG), Belo Horizonte, Brazil*

Abstract

Brazil has a long tradition in the use of health information systems, however – until now – there is no consensus on the minimum data set from which to compose discharge summaries. This article describes the metdodology used by a group of experts – members of WG1 of ISOTC 215 Health Informatics Brazilian mirror committe – to define the information model of the discharge summary. This paper describes the current status of the standardization process and the first pilot tests with this infomation content.

Keywords:
Discharge Summary, Information Model; Continuity of Care; standardization.

Introduction

The hospital discharge summary is an essential document to support efficient and effective continuity of care, allowing for information to be shared [1]. Although Brazil has more than 20 years experience in using health information systems, so far there is no consensus on the minimum data set from which to compose the discharge summary. This article describes the methodology used by a group of experts – members of WG1 of the ISOTC 215 Health Informatics Brazilian mirror committee – to define an information model for discharge summaries.

Materials and Methods

Experts from government, academia, and private institutions – members of WG1: Frameworks, Architectures and Models, from the Brazilian mirror committee of ISO TC 215 Health Informatics – decided to work towards a proposal for a national information model for discharge summaries. A survey was then conducted to gather public and private hospital discharge summary examples around the country, as well as international guidelines (Australian, Canadian, and UK national standards, and ASTM – CCR), and reviews of national and international articles on the subject. [3].

Discharge summaries were collected from different regions of Brazil. One major contribution to the experts' work was a master's degree dissertation from the Catholic University of Paraná that identified the most relevant informational elements used in discharge summary, titled "Identification of the Contents of Standard Discharge Summaries."

Results

The final analysis resulted in a comprehensive discharge summary information model. The full information model is accessible at the WG1 site: http://gt1.medicina.ufmg.br/. The proposal of the discharge summary content was revised by WG1 members at ABNT (an ISO representative in Brazil) to be published for national public consultation in one month.

Conclusion

The forseen schedule entails public consultation ending in February 2015, and the document will be revised and published as a national standard in May 2015, at the latest.

This initiative will certainly contribute to the promotion of better continuity of care in Brazil.

Acknowledgment

The authors express their gratitude to the experts members of the WG1 of ABNT/CEE-78 and to FINEP.

References

[1] Joint Commission International. Padrões de acreditação da Joint Commission International para hospitais, Rio de Janeiro, 2010.
[2] KRIPALANI, Sunil; LEFEVRE, Frank; PHILLIPS, Christopher O; WILLIAMS, Mark V; BASAVIAH, Preetha; BAKER, David W. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA. vol. 297, n.8. Fev. 2007.
[3] ASTM E2369-12, Standard Specification for Continuity of Care Record (CCR), ASTM International, West Conshohocken, PA, 2012. http://www.astm.org/. Available in: (CCR) http://www.astm.org/Standards/E2369.htm. Access: 22 Dez 2014.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-937

The Challenge of e-Health Presence on a Petroleum Platform: Using Telemedicine to Make Operation of Pre-Salt Wells a Reality

R Nunes Ferreira[a], T Lopes da Rosa, C Benevenuto de Campos Lima[b], G Brito Alves de Lima[b], P dos Santos Ramos[c], T Dias da Silva, A Barbieri[a], E Takeo Ueda[a]

[a]Technological Research Institute, IPT, São Paulo, Brazil
[b]Federal Fluminense University, UFF, Niterói, Brazil
[c]Halliburton, Rio de Janeiro, Brazil

Abstract

Telemedicine can be defined as the use of electronic media for the transmission of clinical data and information from one location to another using information technology and telecommunication in order to provide immediate clinical health care at long distances. This new approach can involve specialized medical service centers in the oil production at great distances from the offshore installations in Brazil. The importance of the right health diagnosis, taken at the proper time, will make a serious difference in the facilities, which will be located around 300 km offshore. This paper presents an overview of telemedicine and its different applications, comparing them according to level of maturity and applicability. Important results from a case study in a fixed oil platform are analyzed. At the end of this work, the strategy of telemedicine implementation in a Brazilian petroleum operator is discussed.

Keywords:

MeSH; E-Health; Telemedicine

Introduction

With more than 352 square kilometers, Santos is the largest marine sedimentary basin in Brazil, going from the southern coast of Rio de Janeiro to Santa Catarina State. In this area, there are nine oil and gas production facilities and one gas onshore processing plant. There are two fixed production units and FPSOs (Floating Production Storage and Offloading) [12]. Located more than 300 km off the Brazilian coast, these production units can have more than 100 people on board each unit. In case of emergency situations they must have a specialized medical service. The long distances from land make it impossible to move people rapidly to a hospital, and decisions on the health of patients should be made in the shortest possible time. In this scenario, there is a need to have a structured telemedicine system and partnerships with specialized Brazilian hospitals [1].

Methods

The method proposed in this paper consists of the implementation of a telemedicine pilot project in an oil operator in Santos basin. One important factor to consider is the improvement of assistance to victims of accidents or illnesses on offshore platforms and areas without specialized medical service. The telemedicine system applied to an oil operator will be able to transmit images and sound between different terminals, enabling clinical examination at a distance [2].

Results

The results of the pilot project showed important medical achievements. It was possible to perform diagnosis from the hospital using medical information received from the platform. During the pilot period there were 124 calls in a period of 65 days, with an average of 1.9 visits per day, through connections with videoconferencing.The telemedicine pilot proved the technical feasibility, even using existing communication systems such as satellite networks and radio links. In the first pilot steps it was not clear who should call the hospital's medical specialists, the HSE team onboard or the doctor on duty onshore. The pilot results clarified this. It will be necessary to train all involved about the technologies of videoconferencing and new digital medical equipment.

Conclusion

There was consensus across the medical staff that the technology associated with a large hospital as a partner can support health services, even with an increase in the number of people on board and new production facilities, as is the situation at the studied company. Even with the existing infrastructure of communication it was possible to perform all the services demanded by the pilot, increasing the reliability of offshore diagnosis and improving support to onshore decisions.

References

[1] Cabral J, Kim Y. Multimedia Systems for Telemedicine and Their Communications Requirements. IEEE Communications Magazine 2006;34(7):20-7.

[2] Choi YB, Krause JS, Seo H, Capitan KE, Chung K. Telemedicine in the USA: Standardization through Information Management and Technical Applications. IEEE Communications Magazine 2006;44(4):41-8.

Address for correspondence

Rodrigo Nunes Ferreira (Rodrigo.ferr@uol.com.br)
330, Republica do Chile ave
Zip Code: 20031-170
Rio de Janeiro, RJ, Brazil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-938

Archetype Development Process of Electronic Health Record of Minas Gerais

Thais Abreu Maia[a], Cristiana Fernandes De Muylder[b], Rodrigo Mendonça Queiroga[c]

[a] *Project Manager at Health Department of Minas Gerais, Belo Horizonte, MG, Brazil*
[b] *Professor and researcher of Stricto Sensu Program of Administration and Information System and Knowledge Management, FUMEC University, Belo Horizonte, MG, Brazil*
[c] *Operations Manager at Vivver Systems Ltd., Belo Horizonte, MG, Brazil*

Abstract

The Electronic Health Record (EHR) supports health systems and aims to reduce fragmentation, which will enable continuity of patient care. The paper's main objective is to define the steps, roles and artifacts for an archetype development process (ADP) for the EHR at the Brazilian National Health System (SUS) in the State of Minas Gerais (MG). This study was conducted using qualitative analysis based upon an applied case. It had an exploratory purpose metodologically defined in four stages: literature review; descriptive comparison; proposition of an archetype development process and proof of concept. The proof of concept showed that the proposed ADP ensures the archetype quality and supports the semantic interoperability in SUS to improve clinical safety and the continuity of patient care.

Keywords:

Electronic Health Record; State Medicine; Archetype Development Process.

Introduction

The EHR is defined as one or more repositories integrated that supports continuity of patient care [1]. The health system of Minas Gerais State choose ISO Standard 13606 to develop it. The ADP is a type of knowledge management carried out by health professionals, which ensures consistency of archetypes used in the EHR [1] [2]. However, there is no literature about systematizing archetypes modelling [3]. Considering this scenario, what should be the steps, roles and artifacts for the archetypes development process used in EHR for SUS in MG?

Methods

This is a qualitative study which was divided into 4 stages: 1- literature review; 2- descriptive comparison; 3- archetype development process for SUS in MG and 4- proof of concept.

Results

Stage 1: identified the work of [1], [2], [3], [4] and [5].

Stage 2: systematized steps, roles and artifacts identified.

Stage 3: established ADP for EHR of SUS in MG that followed a process known as PDCA proposed by [2].

Stage 4: approved the model.

Discussion

The proposed ADP allows for control of the archetype development and especially the reduction of time to modeling it to according to SUS legislation. Additionally, the incorporation of management requirements to archetype quality requirements support the reduction of rework.

Conclusion

It is considered that the study met the objective. Study findings will enable increased data reliability for the secondary data use, allow better adaptation of the archetypes of science development speed in the health area and support the semantic interoperability in SUS to improve clinical safety, the continuity of patient care and archetype quality through its governance. Evaluation of necessity to determine time for each step is suggested as a future study.

Acknowledgments

Thanks for support to FAPEMIG, CNPq and CAPES.

References

[1] Santos, MR. Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais 2011. 178 p. Thesis. Belo Horizonte: Escola da Ciência da Informação, Universidade Federal de Minas Gerais, 2011.

[2] Conde, AM. Towards best practice in the Archetype Development Process. 96 p. Dissertation. Dublin: Department of Computer Science, Trinity College of Dublin, 2010.

[3] Späth, MB; Grimson, J. Applying the archetype approach to the database of a biobank information management system. Int J Med Inform 2011 Mar.; 80 (3): 205-226.

[4] Menárgues-Tortosa, M; Fernández-Breis, JT. OWL-based reasoning methods for validating archetypes. J Biomed Inform 2013 Apr; 46 (2): 304-317.

[5] Trigo, JD; Kohl, CD; Eguzkiza, A; Espronceda-Martinez, M; Serrano, AAL; Garcia, J; Knaup, P. On the seamless, harmonized use of ISO/IEEE11073 and openEHR. IEEE J Biomed Health Inform v. 18, n. 3, May 2014.

Address for correspondence

thais.abreu@saude.mg.gov.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-939

Unanticipated consequences of hospital-based insulin management order sets

Eunme Cha, Sherita Hill Golden, Joseph Finkelstein

Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Abstract

Recent studies demonstrated risks to patient safety associated with implementation of electronic applications for medication management in ambulatory care. This study was aimed at demonstrating similar phenomenon in a hospital setting. After introduction of computerized order set targeting hypoglycemia, the frequency of hypoglycemia significantly decreased from 1/1/07 to 12/31/08. In contrast, the frequency of hyperglycemia increased at the same time from 1/1/07 to 12/31/07. Only after subsequent introduction of a hospital-wide standardized insulin order set including hyperglycemia policies frequency of hyperglycemic episodes declined. Hypo/hyperglycemia is associated with adverse clinical outcomes in the inpatient setting. Retroactive analysis showed that if hypoglycemic and hyperglycemic policies were introduced simultaneously, unexpected increase in frequency of hyperglycemic episodes could have been avoided. These data are informative in identifying unanticipated consequences of an insulin management order sets focused entirely on hypoglycemia. A balanced approach in implementing insulin management guidelines concurrently addressing both hypoglycemia and hyperglycemia policies is warranted.

Keywords:

order sets, patient safety, hypoglycemia, hyperglycemia, clinical guidelines

Introduction

In this study, we determined the monthly trend in glycemic control parameters from 1/1/07 to 12/31/08, using the hospital's glucometer database. Throughout 2006 and 2007, a Hypoglycemic Policy, a Diabetes Nursing Superuser Program, and a uniform computerized subcutaneous insulin order set were implemented at Johns Hopkins Hospital. In addition on January 1, 2008, the Hyperglycemia Policy became effective for management of hyperglycemia at our hospital. It included a Hyperglycemia Orderset that guided prescribers in ordering basal, nutritional, and correction insulin with an emphasis on ensuring that insulin deficient patients were to always receive basal insulin and that patients treated with insulin on admission were not to be managed with sliding scale insulin only. We hypothesized that there would be a downward trend in percentage of patients with hypoglycemic and hyperglycemic episodes.

Materials and Methods

Implementation of a hospital-wide Hypoglycemia Policy resulted in a significant reduction in hospital-wide hypoglycemia. However, prior to a complimentary Hyperglycemia Policy, the frequency of hyperglycemia unexpectedly increased. Although not statistically significant, there has been a trend of declining hyperglycemia frequency, particularly severe hyperglycemic episodes (glucose levels >=300 mg/dL), following the implementation of a hospital-wide Hyperglycemia Policy and uniform subcutaneous insulin order set.

Discussion

Our data suggested that simultaneous, as opposed to consecutive, implementation of hypo- and hyperglycemia management strategies may have a more favorable effect on inpatient hyperglycemia trends and overall quality of care. A major limitation of our analysis is that we were unable to distinguish individuals with diabetes from those without diabetes in our existing point of care testing glucose database. While we would assume that the majority of the patients receiving finger-stick glucose readings have diabetes, this may not be the case. In our future analyses, glucometric results have to be stratified by the presence versus absence of diabetes to account for potential confounders in our observed associations.

Conclusion

Introduction of a hospital-wide glucose management program had a significant impact on reducing the frequency of hospital-wide hypoglycemia. Implementation of in-hospital insulin management guidelines may be optimized by introduction of a balanced approach addressing both hypo- and hyperglycemia simultaneously.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-940

The Impact of Implementing a New Computerized Physician Order Entry (CPOE) System on Pharmaceutical Interventions in a Tertiary Brazilian Hospital

Vladimir Ribeiro Pinto Pizzo[a], Paula Brito Gonçalves[b], Lívia Maria Goncalves Barbosa[c], Debora Cecilia Mantovani Faustino de Carvalho[c], Margareth Ortiz de Camargo[d], Carlos Onofre de Lira[e]

[a] *Adult Intensive Care Unit, Hospital Sírio Libanês, Brazil*
[b] *Informatics Technology Division, Hospital Sírio Libanês, Brazil*
[c] *Pharmacy Division, Hospital Sírio Libanês, Brazil*
[d] *CIO, Hospital Sírio Libanês, Brazil*
[e] *Hospital Director, Hospital Sírio Libanês, Brazil*

Abstract

We analyzed trends in pharmaceutical interventions during the implementation of a new computerized physician order entry (CPOE) process in a tertiary hospital in Brazil. The new process utilized an eletronic interface that was designed in-house and an automatic order extension program. The new process reduced the number of order transcriptions and mitigated other potential CPOE-related errors [1].

Keywords:

Computerized Provider Order Entry, Medication error, Intervention.

Introduction

The beginning of our eletronic health system (EHS) implementation process was characterized by interface-related difficulties [2], Due to this, our organization maintained a setting for paper-based order entry with transcriptions and a 24/7 pharmaceutical evaluation for non-staff physicians. When we developed a more intuitive local solution associated with an automatic copy of the order entry ("copy robot") it was possible for physicians to place orders electronically. The aim of this poster is to show the trend of pharmaceutical interventions during this proccess.

Methods

We chose three reasons for order adjustment as study variables: "system problems," "adequacy of pharmaceutical form," and "duplication." These reasons for adjustment were indicated during routine pharmaceutical evaluation of orders. The data were evaluated for two periods–six months before and after June 2014–when the "copy robot" and CPOE systems were activated. The data were plotted in Microsoft Excel and statistically analyzed using an ANOVA single factor test. P values less than 0.05 were considered statistically significant.

Results

The number of prescriptions that needed an intervention decreased during the time periods studied ($p < 0.001$). There was no statistically significant difference in the number of pharmaceutical interventions for "system problems," an increased number for "adequacy of pharmaceutical form" ($p < 0.001$), and a decreased number for "duplication" ($p < 0.001$). Figure 1 shows these results. Light grey columns correspond to "adequacy of pharmaceutical form," dark grey to "duplications," and black to "system problems."

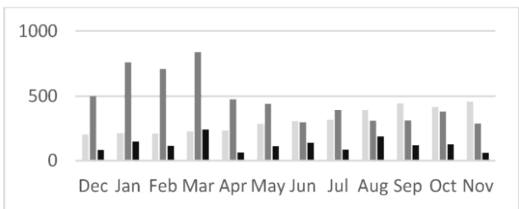


Figure 1 - Pharmaceutical Interventions Over Time

Discussion

Our strategy reduced the number of order transcriptions and minimized some errors like order "duplications." An increase in adjustments for "adequacy of pharmaceutical form" is considered a less dangerous pattern for our population.

Conclusion

The implementation models for electronic health systems should be adapted for each setting and place. Planning and continuous evaluation are key factors for success.

References

[1] Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The Effect of Eletronic Prescribing on Medication Errors and Adverse Drug Events: a Systematic Review. J Am Med Inform Assoc. 2008;15:585-600.

[2] Fumis RRL, Costa ELV, Martins PS, Pizzo VRP, Souza IA and Schettino GPP. Is the ICU staff satisfied with the computerized order entry? A cross-sectional survey study. Rev Bras Ter Int;2014:26(1):1-6.

Address for correspondence

Vladimir Ribeiro Pinto Pizzo, M.D. Hospital Sírio Libanês, 91, Dona Adma Jafet St, São Paulo – SP, Brazil.

vladimir.pizzo@hsl.org.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-941

Methodology to Establish Associations between Data and Clinical Assessment for Computerized Nursing Process in Intensive Care Units

Daniela Couto Carvalho Barra[a], Grace Teresinha Marcon Dal Sasso[a], Fernanda Paese[a]

[a]Nursing Departament, Federal University of Santa Catarina, Brazil.

Abstract

Combining the Information and Communication Technologies (ICT) in the Nursing Process (NP) is a way to support their development in health contexts. The alliance between ICT and the NP integrates and organizes a logical structure of data and clinical information supporting nurses in decision-making. This manuscript describes the methodology used to articulate data and information cynical of Computerized Nursing Process (CNP), according to ICNP® 2.0, associating detailed clinical assessment of each human system to their diagnoses, interventions, and patient outcomes. This is a methodological study and technological production conducted in 2010, and is developed in three stages. It was possible to restructure the CNP from the associations between the data and clinical information of all human systems (cardiovascular, neurological, respiratory, renal, gastrointestinal, cutaneous, musculoskeletal, female / male and biopsychosocial) to their diagnoses, interventions, and results of Nursing.

Keywords:

Nursing informatics; Nursing Process; Terminology; Classification; Intensive Care Unit

Introduction

The Nursing Process (NP) is a care technology that guides the sequence of clinical reasoning and improves the quality of care. Combining the Information and Communication Technologies (ICT) in the NP is the way to support their development in health contexts. The alliance between the ICT and the NP promotes the quality and safety of care because it integrates and organizes a logical structure of data and clinical information supporting nurses in decision-making[1,2] .

The aim of this manuscript is to describe the methodology used for the articulation of data and clinical information of Computerized Nursing Process (CNP) according to ICNP® 2.0 associating detailed clinical assessment of each human system to their diagnoses, interventions and patient outcomes.

Materials and Methods

Methodological and technological production study conducted in 2010 by seven participants, a teacher and six post-graduate course students.

The association of data and clinical information was developed in three steps[2]:

1. Discussion and understanding of the standard International Organization for Standard 18.104. This standard includes the reference terminology model for diagnoses and nursing actions, thus making the standard a tool that facilitates the mapping of various terminologies and promoting the integration of information systems and electronic medical records.
2. Theoretical study on the International Classification for Nursing Practice (ICNP®) 2.0.
3. Combination of data and clinical information to diagnoses, interventions and outcomes of nursing. The base of knowledge was organized according to ICNP® 2.0.

The study was approved by the Ethics of the Federal University of Santa Catarina Committee (protocol 947/10).

Results

It was possible to restructure the CNP from the association between the data and clinical information of all human systems (cardiovascular, neurological, respiratory, renal, gastrointestinal, cutaneous, musculoskeletal, female / male and biopsychosocial) to their diagnoses, interventions and results of Nursing.

The ICNP® version 2.0 reflects major reformulations, as pointed out by nurses, and provides a technology rating system more robust and affordable for these professionals working in Intensive Care Units.

Conclusion

The method adopted from the three steps allows realization of the associations between data and clinical information, diagnoses, interventions, and outcomes of nursing, provides documentation on clinical practice, and support for ICU nurses' decision-making. The ICNP® is used as a basis of knowledge that facilitates organized thinking and clinical judgment for nurses.

References

[1] Barra DCC, Sasso GTMD. Tecnologia móvel à beira do leito: processo de enfermagem informatizado em terapia intensiva a partir da CIPE 1.0. Texto Contexto Enferm. 2010;19(1):54-63.

[2] Sasso GTMD, Barra DCC, Paese F, Almeida SRWA, Rios GC, Marinho MM, Debétio MG. Processo de enfermagem informatizado: metodologia para associação da avaliação clínica, diagnósticos, intervenções e resultados. Rev Esc Enferm USP. 2013; 47(1):242-249.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-942

Nursing Software for Emergency Triage (NSET)

Mandirola Brieux H. F.[abcd], Guillen S. [b], La Rosa F. [c], Moreno C. [c], Benitez S. [d]

[a] *Belgrano Hospital in San Martín, Buenos Aires, Argentina*
[b] *BIOCOM, Buenos Aires, Argentina*
[c] *HL7 Argentina, Buenos Aires, Argentina*
[d] *Hospital Italiano, Buenos Aires, Argentina*

Abstract

Determining the priority of attention in an Emergency Room (ER) has always been a difficult issue. Priority is determined with a simple triage system as people arrive at the hospital. It is important to establish how long they can wait for treatment. In order to obtain the best assessment of patients' conditions, we built a Nursing Software for Emergency Triage (NSET). The objective of this work was to assess the efficacy of the NSET versus the triage process without any software (TWS). Results showed that the NSET we built was a substantial help. With this software, we decreased significantly:1) the length of the triage system process, 2) the waiting time of patients in the waiting room, 3) the number of complaints and 4) the number of patients who walk away. In conclusion, the NSET improves and helps to define more accurately a patient's risk. NSET helps in the emergency department triage.

Keywords:

EMR; HL7;TRIAGE; Nurse; Computer; System; eHealth; MTS-Manchester; Medical Informatics; ER.

Introduction

In the Emergency Room (ER) staff do not attend to patients on a first come, first served basis; but according to the severity of the patient's condition. In an ER, we *must* know who needs help more urgently in the waiting room. Patients presenting themselves in the ER currently face unacceptable delays in initial treatment, and long, costly hospital stays due to suboptimal initial triage and site-of-care decisions. For the purpose of this work, the Manchester Triage System (MTS) was used to build a Nursing Software for Emergency Triage (NSET) to be used by triage nurses. Our hypothesis is that the NSET is better than a triage system without software (TWS) to categorize patients at the ER. [1,2]

Materials and Methods

Belgrano Hospital is located in the surrounding area of Buenos Aires city, Argentina, South America. It is a medium-level acute hospital. There are 200 beds, 8 critical care unit beds and 250 physicians. This study is a prospective study. The period of data collection for this work was from April 30th, 2014 to November 1st, 2014. In this period, we collected 1800 patients' data in the ER of the Belgrano Hospital, Buenos Aires, Argentina. During the data collection period we included patients who were treated in the ER. Twenty eight patients in total were excluded from the sample because some of their data were missing. The final sample included 1772 patients. Fifteen of them belonged to the NEST group and thirteen to the control group.

Results

The results were as follows: The triage time in the control group was 3.7 and in the NSET group 2.5 with a p less than 0,001. For the Left-without-being-seen index, the results for the control group were 4. 76 and for the NSET group 3. 98. The difference between the two samples was not significant waiting time of patients in the ER waiting room. In this case, there was a huge difference between both samples; the control group with an average time of 97 minutes against the NSET group with just 28 minutes. In the Complaint-index, the difference between the two groups was significant with the control group at 42.01 and the NSET at 14.62.

Discussion

Results showed that the NSET was of significant help and was better than the TWS. NSET allowed solving several problems and resulted in a better organization, whereas TWS did not.

Conclusion

Treatment priority in the ER may be subject to variation due to different levels of patient illness. We must give more importance to determining the risk for each patient in the moment they enter in the ER which is why triage is critical. The NSET improves and helps to assess more accurately the condition of a patient.

References

[1] Schuetz P, Hausfater P, Amin D, Haubitz S, Fässler L, Grolimund E, et al. Optimizing triage and hospitalization in adult general medical emergency patients: the triage project. BMC Emerg Med. 2013;13(1):12.

[2] Van Veen M, Steyerberg EW, Ruige M, van Meurs AHJ, Roukema J, van der Lei J, et al. Manchester triage system in paediatric emergency care: prospective observational study. BMJ. 2008;337:a1501.

Address for correspondence

humberto.mandirola@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-943

Nursing Clinical Documentation System Structured on NANDA-I, NOC, and NIC Classification Systems[1]

Heloisa Helena C. Peres[a], Diná de Almeida Lopes M. da Cruz[a], Michelle Tellez[b], Rita de Cassia Gengo e Silva[c], Regina Celia dos S. Diogo[d] , Diley Cardoso F. Ortiz[e], Dóris R. Ortiz[f]

[a] PhD, RN, Professor, School of Nursing, University of Sao Paulo, Brazil; [b] PhD, RN, Associate Professor, Department of Nursing and Health Sciences, California State University East Bay, USA; [c] PhD, RN, Assistant Professor, School of Nursing, University of Sao Paulo, Brazil; [d] PhD, RN, Candidate, School of Nursing, University of Sao Paulo, Brazil; [e] MSN, RN, University Hospital, University of São Paulo, Brazil; [f] Undergraduate Nursing Student, School of Nursing, University of Sao Paulo, Brazil

Abstract

Information is a key feature that health professionals need to exercise their profession with efficiency and quality. This study aims to present the experience of the usage of an electronic system for clinical documentation in nursing in a university hospital. It is a methodological research of technology production. The system was developed in four phases: Conception, Elaboration, Construction, and Transition, and was named Electronic Documentation System of the University of São Paulo Nursing Process (PROCEnf-USP™). The knowledge base of PROCEnf-USP™ was organized in hierarchy of domains and classes, according to NNN linkages.

Keywords:

Nursing Information Systems; Information Systems; Nursing Diagnoses; Classification System; Nursing Process.

Introduction

Health records are the primary information tool to improve clinical communication and evidence based practice. Structured documentation produces data that generate more meaningful and reliable information than free documentation [1]. Classification systems are instruments to improve the reliability, validity and usability of the nursing documentation [2]. The implementation of the NNN linkages (classifications of nursing diagnoses, outcomes and interventions) [2] in electronic nursing documentation system encourages nurses to adopt the nursing process, proves the diagnostic accuracy and the scope of the results obtained from the patients**.** The objective of this study is to present the experience of the usage of an electronic system for clinical documentation in nursing in a university hospital.

Methods

In this methodological research four cyclical phases of creation and assessment of technological product were used to develop the system, based on project management model in the Project Management Institute (PMI) [3]. Phases: Conceptualization - project scope was approved and defined. Estimated resources, problems and expected benefits were identified. Elaboration: problem domain was analyzed in order to complement the survey documentation of the use cases and modeling of system data. Construction: system development (prototype), aiming to refine the requirements, build and test their components. Transition: dissemination of the new system for the whole nursing team, user training, and evaluation of user satisfaction [3].

Results

The system developed is Electronic Documentation System of the University of São Paulo Nursing Process (PROCEnf-USP™). It allows clinical documentation, generates reports of the nursing process and provides decision support about diagnoses, expected results and nursing interventions [3]. This support system for clinical decision has both professional and academic environments, and it has been used since 2009 [1].

Conclusion

This project was funded by Brazilian government agencies for research support. PROCEnf-USP™ generates data for scientific research regarding the evaluation, accuracy and usability of the system.

References

[1] Maia FOM, Ortiz DCF, Oliveira NB, Trindade MM, Gomes AVS, Pimenta CAM. Sistema informatizado de apoio à decisão diagnóstica: a experiência do Hospital Universitário da Universidade de São Paulo. In: Herdman TH, Lopes MVO, Almeida MA, Chianca TCM. PRONANDA. 2nd cycle. Porto Alegre: Artmed, 2014.

[2] Jonhson M et al. Ligações entre NANDA, NOC e NIC - Diagnósticos, Resultados e Intervenções de Enfermagem. Trad. De Regina Machado Garcez. 2ª ed. Porto Alegre: Artmed, 2009.

[3] Peres HHC, Cruz DALM, Lima AFC, et al. Desenvolvimento de Sistema Eletrônico de Documentação Clínica de Enfermagem estruturado em diagnósticos, resultados e intervenções. Rev. esc. enferm. USP 2009: 43(2):1149-55

Address for correspondence

Heloisa Helena C. Peres, Professor, School of Nursing, University of Sao Paulo. Av Dr Eneas de Carvalho Aguiar, 419 – 1º andar Cerqueira Cesar – São Paulo CEP 05403-000 Telefone: + 55 (11) 30617552/ E-mail: hhcperes@usp.br

[1] National Council for Scientific and Technological Development (CNPq)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-944

A Trial of Nursing Cost Accounting using Nursing Practice Data on a Hospital Information System

Akiko Miyahira[ab], Kazuko Tada[ab] Masatoshi Ishima[ab], Hidenori Nagao[a], Tadashi Miyamoto[a], Yoshiaki Nakagawa[c], Tadamasa Takemura[b]

[a] *Hyogo Prefectural Amagasaki Hospital, Amagasaki, Japan*
[b] *Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan*
[c] *Shikoku Medical Center for Children and Adults, Zentsuji, Japan*

Abstract

Hospital administration is very important and many hospitals carry out activity-based costing under comprehensive medicine. However, nursing cost is unclear, because nursing practice is expanding both quantitatively and qualitatively and it is difficult to grasp all nursing practices, and nursing cost is calculated in many cases comprehensively. On the other hand, a nursing information system (NIS) is implemented in many hospitals in Japan and we are beginning to get nursing practical data. In this paper, we propose a nursing cost accounting model and we simulate a cost by nursing contribution using NIS data.

Keywords:
Nursing Cost Accouting; Hospital Administration; Nursing Information System.

Introduction

In many cases, the results of nursing cost accounting are unrealistic for nursing staffs. Part of reason that it is difficult to adapt to activity-based costing is because it is hard to figure out realistic nursing activities [1]. A nursing information system (NIS) includes nursing practical data. So we consider that we are able to make a new nursing cost accounting model using these data.

Methods

Nursing cost accounting model

We are able to categorize nursing practices into three tasks: (a) variable tasks to patients, (b) fixed tasks to patients, and (c) other fixed tasks.

1. Variable tasks to patients include individualized care for each patient. For example, tasks such as toilet support, bed bath, and so on.

2. Fixed tasks to patients include basic infomaiton collection, orientation of administration, and so on. These tasks are performed regardless of patient.

3. Other fixed tasks are education, operating report, and so on. All nurses have the job of administration in the ward.

These tasks have some indicators of working. They are hours, payment, skill levels and so on. Therefore we can calculate costs about these indicators.

Data Extraction from NIS

Hyogo Prefectural Amagasaki Hospital has implemented a full EMR system. The EMR system has full NIS and the NIS has nursing practical data for each patient. Hyogo Prefectural Amagasaki Hospital has 192 categories of nursing practice and NIS has all nursing practical data of these categories.

Results and Discussions

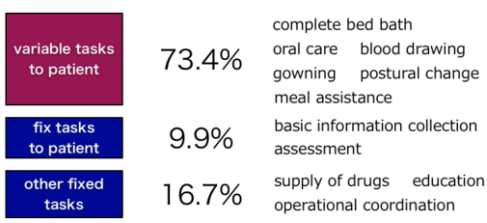


Figure1 – The ratio of three tasks

We could simulate the amount of nursing practices in the ward. For example, we analyzed variable tasks to patient by Diagnosis Procedure Combination (Japanese DRG) and we could predict nursing tasks and nursing costs.

References

[1] K. Tanaka et al. A simulation model of hospital management based on cost accounting analysis according to Disease, Journal of Medical Systems, Vol 28 No. 6, p689-709, 2004.

Address for correspondence

Tadamasa Takemura, Graduate School of Applied Informatics, University of Hyogo, Minatojima-Minamimachi 7-1-28, Chuo-ku, Kobe, Japan.

Email: takemura@ai.u-hyogo.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-945

An Electronic Nursing Patient Care Plan Helps in Clinical Decision Support

Wong CM[a], Wu SY[a], Ting WH[b], Ho KH[b], Tong LH[a], Cheung NT[a,b]

[a] *Health Informatics Section, Hospital Authority, Hong Kong SAR*
[b] *Information Technology Division, Hospital Authority, Hong Kong SAR*

Abstract

Information technology can help to improve health care delivery. The utilisation of informatics principle enhances the quality of nursing practices through improved communication, documentation and efficiency. The Nursing Profession constitutes 34% of the total workforce in the Hong Kong Hospital Authority (HA) and includes 21,000 nurses in 2012. To enhance the quality of care and patient safety in both hospitals and community care setting, it is essential that an integrated electronic decision support system for nurses is designed to track documentation and support care or service including observations, decisions, actions and outcomes throughout the care process at each point-of-care. The Patient Care Plan project was set up to achieve these objectives. The Project adheres to strict documentation information architecture to ensure data sharing is freely available. Preliminary results showed very promising improvement in clinical care.

Keywords:

Patient Care Plan; Efficiency; Documentation.

Introduction

The Patient Care Plan Project (PCP) is a new nursing function in the HA Clinical Management System (CMS) that depicts the identified care needs, treatment goals and the progress towards meeting the goals. Clear communication and appropriate assessment form the basis of providing vital and essential information for nursing interventions and initiating patient care plan in daily operation.

Methods

The HA information architecture was developed in 2002 to support a fast, robust and standardised electronic patient record (ePR) from data captured from various CMS function modules. The Generic Clinical Documentation (GCD) engine would manage the lifecycle of various information architecture elements such that the semantics of these elements (regardless of their source, whether internal or external) are preserved from design, creation, capture, use, reuse, and analysis. Forms created using this process and architecture will be semantic interoperable in different service modules. The PCP is constructed using this architecture.

The PCP is designed to indicate and identifies high risk patients based on assessments to provide early notifications to medical officers, nurses and allied health professionals. Validation rules and auto-calculation were utilised to assist and support decisions. System alert and logic are built-in and documented to track ongoing assessments, actions and responses to the high risk indicators. Risk prevention outcomes and standards from the current Nursing Standards for Patient Care in HA have been hard-coded in the system.

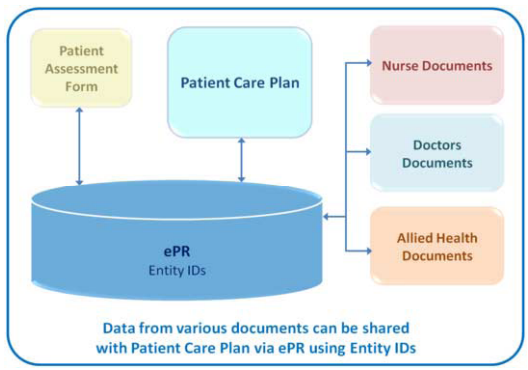


Figure 1 - Data sharing to support patient care plan

The project is governed by the Patient Care Plan Working Group under the Clinical Informatics Program Steering Group. An overview of the PCP has been included in the summary page and the care plan will be shared in the ePR for future development. Agreement for complete data sharing by members in the nursing systems starting from admission to hospitalisation and discharge with follow-up in outpatient department and community setting will achieve the requirements of standardisation of entities and terms (figure 1).

Results

The PCP helps to alert HA healthcare professionals to implement relevant, appropriate and specific interventions for high-risk patients. Decisions are based on professional knowledge supported by a comprehensive and coherent assessment which is freely available in the electronic medical record system. This co-ordinated flow of structured data and care planning processes are essential to all stakeholders in the patient's journey to maximise benefits for the patient.

Conclusion

The information architecture of the PCP excels itself in a clinically meaningful way; not only it enables data sharing from admission to discharge, data can be reused across different CMS function modules, and alerts HA healthcare professionals to response to high risk patients in risk reduction and prevention. There is pronounced added value in the Patient Care when end users are prompted to offer timely and appropriate care support to patients.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-946

Clinical Decision Support to Implement CYP2D6 Drug-Gene Interaction

Pedro J. Caraballo, Mark Parkulo, David Blair, Michelle Elliott, Cloann Schultz, Joseph Sutton, Padma Rao, Jamie Bruflat, Robert Bleimeyer, John Crooks, Donald Gabrielson, Wayne Nicholson, Carolyn Rohrer Vitek, Kelly Wix, Suzette J. Bielinski, Jyotishman Pathak, Iftikhar Kullo

Mayo Clinic, Rochester, Minnesota, USA

Abstract

The level of CYP2D6 metabolic activity can be predicted by pharmacogenomic testing, and concomitant use of clinical decision support has the potential to prevent adverse effects from those drugs metabolized by this enzyme. Our initial findings after implementation of clinical decision support alerts integrated in the electronic health records suggest high feasibility, but also identify important challenges.

Keywords:

Clinical Decision Support Systems; Pharmacogenomics; Cytochrome P-450 CYP2D6; Electronic Health Records.

Introduction

Codeine and tramadol are opioid prodrugs commonly used in clinical practice to treat moderate pain. They have been associated with several adverse effects including respiratory depression and death. They are metabolized in the liver by the cytochrome P450 2D6 enzyme and transformed to their active metabolites which are responsible for their therapeutic effects. CYP2D6 enzyme is encoded by the *CYP2D6* gene and the enzyme activity varies among individuals but this variation can be predicted by pharmacogenomic (PGx) testing. In the case of codeine, high activity of CYP2D6 produces high levels of morphine and increases the risk of adverse effects. On the other hand, low activity produces low levels of morphine resulting in less pain control.

We present our initial findings after the implementation of PGx clinical decision support (CDS) rules integrated in the electronic health records (EHR) to alert on actionable polymorphism of *CYP2D6* gene to optimize dosing and prevent adverse effects from codeine and tramadol.

Methods

We developed a series of CDS rules to alert providers based on the results of PGx testing for *CYP2D6* and the documentation of the genotypes and phenotypes in the EHR. There were approximately 640 different genotype variations that were matched to eight phenotypes based on the level of enzyme activity: high metabolic activity (ultrarapid, extensive to ultrarapid, intermediate to ultrarapid), relatively normal activity (extensive, intermediate to extensive, intermediate) and low activity (poor to intermediate, poor). The CDS rules displayed alerts only for high or low activity. The CDS rules assessed all new *CYP2D6* genotype test results, matched them to one of the eight phenotypes, and documented the select phenotype as a problem in the problem list. Other rules assessed all codeine/tramadol orders and reviewed the problem list and PGx results. If there was an actionable result, the rule displayed a CPOE (Computerized Physician Order Entry) screen rule. The system also detected unreadable results and sent an email to the implementation team to initiate a manual review and documentation. All the alerts had a clear message indicating the clinical impact of the specific drug-gene interaction. They also had links to additional PGx education and treatment alternatives.

Results

Between 10/27/2013 and 7/31/2014 a total of 206 events were triggered by the CPOE screen rule and 45 (21.8%) of them were unreadable. The main cause of unreadable result was lack of discrete values on the PGx reports. The other 161 events displayed alerts due to increased activity (121, 75%) or decreased activity (40, 25%), see Table 1. The number of events was higher than the number of patients and providers, indicating that the same alert displayed multiple times for the same drug/patient/provider interaction.

Table 1 – CDS Events Based on CYP2D6 Phenotypes

Phenotypes	Events (n)	Patients (n, female %)	Providers (n)
Ultrarapid	49	18, 67%	27
Extensive to ultrarapid	72	24, 46%	39
Intermediate to ultrarapid	0	0	0
Poor to intermediate	15	12, 42%	14
Poor	25	13, 69%	17
Unreadable	45		

Conclusion

Implementation of CDS integrated in the EHR to selectively prevent CYP2D6 related drug-gene interactions is feasible, but encounters significant challenges. The PGx Lab reports are still evolving and they are in need of additional standardization to support stable electronic data integration with the EHR and to facilitate execution of CDS interventions. Provider understanding of drug-gene alerts and immediate response to the alerts requires additional analysis, seeking to prevent unnecessary and repetitive alerts.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-947

A Scalable Architecture for Rule Engine Based Clinical Decision Support Systems

Soumi Chattopadhyay[1], Ansuman Banerjee[1], Nilanjan Banerjee[2]

[1]Indian Statistical Institute, Kolkata, India
[2]IBM Research, India

Abstract

Clinical Decision Support systems (CDSS) have reached a fair level of sophistication and have emerged as the popular system of choice for their aid in clinical decision making. These decision support systems are based on rule engines navigate through a repertoire of clinical rules and multitudes of facts to assist a clinical expert to decide on the set of actuations in response to a medical situation. In this paper, we present the design of a scalable architecture for a rule engine based clinical decision system.

Keywords:

Clinical decision support, Rule engine

Introduction

The core component of CDSS [1] consists of a rule inferencing system where characteristics of an individual patient are matched to a clinical knowledge base, and then patient-specific assessments or recommendations are presented to the clinician(s) and/or the patient for a decision. Commercial rule engines such as Drools [2] and ILOG [3] have shown to satisfy most needs of a CDSS. An evaluation of the performance of ILOG showed that the overall performance was generally satisfactory for small rule sets. However, hundreds of rule sets are possible in a real-time surveillence system and rule engines do not perform very well on large rule sets. In this paper, we propose an efficient and scalable framework for a rule engine based CDSS.

The Proposed Framework

The key idea behind the framework is a cache-based lazy loading mechanism that consists of rule clustering and hashing kernel, coupled with a prediction based technique for rule evaluation and faster actuation. The main motivation for rule clustering is to save the memory requirement for rule loading and to expedite rule firing by the rule prediction strategy. If we can efficiently and accurately predict the set of rules that are expected to be evaluated to truth based on the arriving facts, we can improve the rule firing time by triggering the actuations beforehand. The rule clustering step is done during preprocessing, whereas prediction is done at runtime. In the clustering step, two rules that have a common fact variable in their antecedents are put in the same cluster. The clusters are mutually disjoint. This ensures that only the rules in one cluster are affected by a fact and only that cluster gets loaded into production memory based on the matching facts at once. The prediction mechanism at run time only stores the last evaluation outcome of a rule in history. In this work, we experiment with two different prediction schemes: a deterministic scheme, in which a rule R is predicted to be true if it was true in the previous session, and a probabilistic scheme, in which the prediction is done with a certain probability.

Experimental Results

We implemented the framework with *JRuleEngine* [4] and evaluated its performance with a dataset of 16 rules defined over 32 fact variables. The rule base contains 5 clusters. Performance evaluation experiments were done using randomly generated sets of facts. Table 1 shows the memory gain (in %), the amount of memory saved using lazy loading of rules and used more intelligently for larger rule bases, with lazy loading of rules.

Table 1. No. of facts vs memory gain (in %)

Cardinality of Facts	*1*	*5*	*10*	*15*	*30*
Memory Gain(%)	*87.5*	*18.75*	*5.56*	*3.94*	*1.125*

Figure 1 shows the actuation time required in normal mode of operation, i.e., without prediction and with different prediction algorithms discussed above, against the number of facts. Our prediction based approaches for rule evaluation gains consistently over the normal mode of operation.

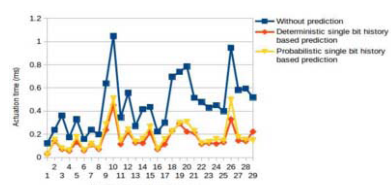


Figure 1. Number of facts vs actuation time

Conclusion

Despite all the benefits of rule engine based CDSS, they suffer from some performance bottlenecks. We propose a framework for improving the overall performance of the rule engine in terms of memory utilization and execution time. Experimental results show promising performance gain that can be leveraged for large-scale operations. In future, we plan to take up the evaluation of our proposed architecture for industry grade performance benchmarks.

References

[1] Berlin et al., A taxonomic description of computer-based clinical decision support systems, Journal of Biomedical Informatics, 2006.

[2] DROOLS: http://www.drools.org/

[3] ILOG: http://www-01.ibm.com/software/info/ilog/

[4] JRULEENGINE: http://jruleengine.sourceforge.net/

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-948

Clinical Decision Support Based on Integrated Patient Models: A Vision

Kerstin Denecke[a]

[a] *Innovation Center Computer Assisted Surgery, University of Leipzig, Germany*

Abstract

Clinical decision making is non-trivial given the amounts of data and knowledge that needs to be considered. So far, medical knowledge, biological knowledge and patient data are separated from each other and need to be integrated mentally by a physician to form an overarching patient model. In this paper, we describe a vision for future decision support systems that link knowledge about organ functions, biological processes, treatment decisions and clinical data represented in repsective models. Requirements and challenges for realizing this vision will be collected.

Keywords:

Clinical decision support system; Digital patient model; Medical informatics applications; Health information system.

Introduction

The challenges we face nowadays in clinical settings, are that clinical data is distributed in several systems and available in multiple formats. Best practices and information about treatment plans and options are captured in biomedical literature or clinical practice guidelines. Not only the access to this data during diagnosis and treatment is a challenge, also its manual or automatic integration, analysis and interpretation. To address these problems, we consider the concept of model-based decision support. A digital patient model integrates various information entities that need to be considered within therapy planning and clinical decision making. It further describes relations between clinical parameters and sociological factors related to a specific clinical pathology. The medical knowledge can be modelled in different ways, and reflects different aspects, e.g. as functional, geometric or probabilistic graphical model. Our vision of model-based decision support is to integrate these different models in a clinical decision support system. Such integration will lead to an evidence- and model-based decision making. It is still unclear, how to integrate these models. In this work, we will describe the general idea and collect the requirements as well as challenges to be addressed for realising the vision.

Methods

Our concept of model-based decision support foresees three types of models: (1) A **decision model or predictive model** (e.g. realized as probabilistic graph) encompasses the medical knowledge on the diagnostic of the disease, the therapy decision including the therapy options, their risks for the patient and their relations to single information items relevant in decision making. (2) A **geometric or functional model** models an organ or organ function, and (3) A **patient specific data model** includes and integrates the relevant patient data (labor tests, examination results, medical report, signal and image data) in a standardized format.

Results

In order to realize a three layer model-based decision support technologies in three main areas are necessary: (1) Modelling of processes, diseases and clinical decision making, (2) Methods for linking data and models, (3) Ontologies and classification systems for describing the concepts of the models and providing the relations for inferencing. From the requirements result the following challenges:

- Modelling the decision processes in form of a digital patient information model is still time-consuming. To ensure a high quality of the models, human interaction is necessary in the modelling process. The challenge is to find the right balance between automatic model generation through data and text mining (at least for probabilistic graphical models) and manual generation. Geometric modelling requires segmentation which can be partially done automatically so far. The challenge is to improve automatic segmentation.
- Collecting all relevant patient data from the multiple information systems (situation description) is difficult due to vendor-lock and a broad landscape of information systems in use.
- Linking data semantically requires corresponding methods for data linking and semantic annotations.
- Missing common data models and use of different vocabularies complicate the integration of domain theory (model) and situation description (real patient data).

Conclusions

The main benefit of the presented approach is that functionalities of current information or clinical decision support systems are extended by:

- bridging the gap between system biological models and patient and clinical data (patient model) as well as medical knowledge data (decision model) – in order to make justified, well-grounded therapeutic decisions, and by
- improving semantic annotation and linking of the heterogeneous data – in order to support users with an integrated view on patient data and on knowledge about disease development and treatment.

Acknowledgements

This research is part of the project Digital Patient and Process Model funded by the German Ministry of Research and Education.

Address for correspondence
Kerstin Denecke / Innovation Center Computer Assisted Surgery

Semmelweisstr. 14, D - 04103 Leipzig, Germany

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-949

Impact of Specific Alerts in Potassium-Increasing Drug-Drug Interactions

Emmanuel Eschmann, Patrick E. Beeler, Jürg Blaser

Research Centre for Medical Informatics, University Hospital, Zurich, Switzerland

Abstract

Alerts in potassium(K^+)-increasing drug-drug interactions (DDIs) are often ignored due to their low specificity. Although different approaches have been implemented to address DDIs, subsequent clinical studies revealed poor adherence to such alerts. We therefore suggest a novel alert concept currently being evaluated in a randomized clinical trial in a large teaching hospital. Highly specific reminders (to monitor K^+) and alerts (of hyperkalaemia) are displayed to the physicians of the intervention group, whereas reminders and alerts are suppressed in the control group. Preliminary analysis shows a high alert specificity. Furthermore, the physicians of the intervention group reacted significantly faster to a problematic situation arising during a K^+-increasing DDI compared to the physicians of the control group, indicating that this concept has an impact on physician behaviour.

Keywords

Decision Support Systems, Clinical; Drug Interactions; Hyperkalaemia; Randomized Controlled Trial

Introduction

Alerts in the context of potassium(K^+)-increasing drug-drug interactions (DDIs) are often ignored due to their low specificity, thus inducing alert fatigue. Approaches addressing this issue include (a) displaying relevant laboratory data in alerts, (b) suppressing less significant alerts by taking into account patient data, (c) removing alerts when the initiating conditions are no longer met, and (d) focussing on high-priority DDIs.

However, the clinical studies examining the impact of these approaches showed poor adherence to DDI alerts. Therefore, we combined these approaches, determined the optimal thresholds by analysing patient data, and are now evaluating the impact of this novel alert concept in a clinical trial.

Methods

A 12-month cluster randomized trial including all inpatients of a large teaching hospital (37,000 inpatients per year) was started in January 2014. Fourteen clinics were randomized to the intervention group and 15 to the control group. ICU stays were excluded. Both, at onset and during a K^+-increasing DDI, the physicians of the intervention group were (a) reminded to monitor K^+, if the current serum K^+ is unknown and no K^+ level measurement has been ordered, (b) warned against hyperkalaemia risk (when serum K^+ ≥4.9mEq/l), and (c) alerted of hyperkalaemia (when serum K^+ ≥5.5mEq/l). All reminders and alerts were non-interruptive, displayed in the electronic health records of the intervention group only, and were removed when the triggering conditions were no longer met. The adherence of physicians to reminders and alerts was reflected by the time elapsed between the reminder or alert activation and its deactivation due to a modification of the K^+-increasing DDI or ordering of a new serum K^+ measurement.

The Mann-Whitney-Wilcoxon test was used for statistics. A *p*-value of ≤0.05 was considered to be statistically significant.

Results

We analysed 1,964 alerts and reminders triggered in the first 10 months of the trial (1,262 displayed in the intervention group, 702 suppressed in the control group). On average 32.2 potentially severe K^+-increasing DDIs occurred per day in the intervention group. However, the physicians of the intervention group were only exposed to 2.8 reminders for K^+ monitoring, 1.1 warnings against increased hyperkalaemia risk, and 0.2 alerts of hyperkalaemia per day.

The physicians of the intervention group reacted significantly faster to an inadequate K^+ monitoring and to an elevated serum K^+ level compared to the physicians of the control group (Table 1).

Table 1 – Reaction time following reminders and alerts

Alert / Reminder	Reason for termination	Intervention group[a]	Control group[a,b]	p
K^+ monitoring reminder	K^+ monitoring	11.2 h	18.0 h	0.04
Alert of elevated K^+ or hyperkalaemia	K^+ monitoring	8.0 h	16.0 h	0.01
	K^+ increasing DDI stopped	4.0 h	6.0 h	0.05

[a] median; [b] alert / reminder displayed, but not logged

Conclusion

This preliminary analysis of our ongoing cluster randomized trial underlines the high specificity of the implemented concept (4.1 reminders and alerts compared to 32.2 K^+-increasing DDIs per day). Moreover, reaction time following reminders and alerts were shorter in the intervention group compared to the control group, i.e. physicians of the intervention group reacted significantly faster to a problem arising during a K^+-increasing DDI compared to the physicians of the control group. Therefore, this novel alert concept has an impact on physician behaviour and may increase patient safety.

The clinical outcome parameters will be analysed after the completion of the study in January 2015 to be presented at MEDINFO2015.

Address for correspondence

Emmanuel Eschmann; University Hospital Zurich
Research Centre for Medical Informatics
Raemistrasse 100; CH - 8091 Zurich; Switzerland
emmanuel.eschmann@usz.ch

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-950

Detecting, Monitoring, and Reporting Possible Adverse Drug Events Using an Arden-Syntax-based Rule Engine

Karsten Fehre[a], Manuela Plössnig[b], Jochen Schuler[c], Christina Hofer-Dückelmann[d], Andrea Rappelsberger[e], Klaus-Peter Adlassnig[a,e,*]

[a] *Medexter Healthcare GmbH, Vienna, Austria*
[b] *Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria*
[c] *Paracelsus Medical University Salzburg – Private Foundation, Salzburg, Austria*
[d] *Landesapotheke am St. Johanns-Spital, Salzburg, Austria*
[e] *Section for Medical Expert and Knowledge-Based Systems, CeMSIIS, Medical University of Vienna, Vienna, Austria*

Abstract

The detection of adverse drug events (ADEs) is an important aspect of improving patient safety. The iMedication system employs predefined triggers associated with significant events in a patient's clinical data to automatically detect possible ADEs. We defined four clinically relevant conditions: hyperkalemia, hyponatremia, renal failure, and over-anticoagulation. These are some of the most relevant ADEs in internal medical and geriatric wards. For each patient, ADE risk scores for all four situations are calculated, compared against a threshold, and judged to be monitored, or reported. A ward-based cockpit view summarizes the results.

Keywords:

Adverse drug events;Arden Syntax rule engine;clinical decision support;drug monitoring; pharmacovigilance.

Introduction

The increasing use of medical drugs has raised the risk of drug-related damage, especially in elderly patients. Although a legal obligation to report adverse drug events (ADEs) has been instituted in several countries, the number of reported cases remains low. Only about 10–20% of medication errors and 1–13% of detected ADEs are reported. ADE detection and reporting is a time-consuming and expensive task. Hospitals need an efficient way to quantify the numbers and severity of ADEs before corrective action can be taken by pharmacists and physicians.

Methods

An ADE cockpit for detecting, monitoring, and reporting possible ADEs has been developed in the iMedication project [1]. The core of the software implementation is a decision support system employing a hybrid approach, combining the IHI Global Trigger Tool method and Morimoto's classification to detect suspected ADEs. The system's knowledge base consists of medical logic modules which encode medical expert knowledge in Arden Syntax [2], and the modules are executed by an Arden Syntax rule engine on a server. The data to be processed are derived from various sources—the hospital information system, an electronic health record, and entered information—and divided into six categories: demographic data, laboratory test results, the patient's symptoms, diagnoses, medications, and hospital events. The aggregated information of a single patient is delivered to the Arden Syntax server, which returns a detailed interpretative summary for each identified ADE consisting of a) an ADE risk score which reflects the degree and severity of the ADE, b) the institutions to be informed, depending on the severity of the ADE, c) the triggers having fired, and d) the complete patient information used for interpretation.

Results

Four clinically relevant situations (hyperkalemia, hyponatremia, renal failure, and over-anticoagulation) were selected as exemplary conditions. The corresponding knowledge bases containing the respective ADE triggers together with their significance, expressed as scores between 1 and 3, were established in close collaboration between experienced clinicians and medical knowledge engineers. All four possible ADEs were analyzed for 100 patients. The test yielded a sensitivity of 88% (3 false negatives out of 25) and a specificity of 87.7% (46 false positives out of 375).

Conclusion

The proposed system is suitable for: (a) quality assurance by retrospective evaluation of clinical data in regard of suspected ADEs, (b) active feedback for clinicians during patient treatment, and (c) pharmacovigilance reporting.

References

[1] Plößnig M, Schuler J, Hansbauer B, Stroka S, Engel T, Hofer-Dückelmann C, Radulescu M, Adlassnig KP. Semi-automated identification of potential adverse drug events. In: Schreier G, Hayn D, Hörbst A, Ammenwerth E, eds. Tagungsband der eHealth2012. Wien: OCG, 2012; pp. 51-6.

[2] Samwald M, Fehre K, De Bruin J, Adlassnig KP. The Arden Syntax standard for clinical decision support: Experiences and directions. J Biomed Inform 2010: 45(4): 711-8.

* Correspondence to klaus-peter.adlassnig@meduniwien.ac.at.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-951

Systems Medicine for Multiple Myeloma: A Review on Decision Support Systems

Matthias Ganzinger, Christian Haux, Christian Karmen, Thomas Wetter, Petra Knaup

Heidelberg University, Institute of Medical Biometry and Informatics, Heidelberg, Germany

Abstract

Systems medicine is a current approach trying to improve treatment for patients with complex diseases by analyzing as much phenotype and genotype data as possible for the disease in question. For individualized treatment decisions in clinical practice, this task has to be supported by an application system with decision support component. For a research project on systems medicine we reviewed methods for decision support. Criteria for selecting a method are derived from characteristics of the data and the diseases. They include, among others: dimensionality of data and existence of a priori models for diseases. As a result we decided to implement a prototype system with a case-based reasoning component for systems medicine on multiple myeloma.

Keywords:
Systems Medicine, Clinical Decision Support Systems, Multiple Myeloma, Case-based Reasoning

Introduction

For complex diseases like cancer, it is known that frequently only a fraction of patients will benefit from a treatment and some will experience more side effects than others. Factors causing response to therapy or side effects are hardly known.

In systems medicine, these problems are addressed by using as many data sources as possible for the treatment decision. This does not only include data on an individual patient, but also data on other patients suffering from the same disease. Data types include genotype data, as they are measured by microarrays and next-generation sequencing, as well as phenotype data. However, to draw clinically relevant conclusions from such a huge amount of data, support through information technology (IT) is inevitable. A decision support component is a key element of an IT system for systems medicine.

Currently, we have built an IT system for the systems medicine research project: "Clinically-applicable, omics-based assessment of survival, side effects, and targets in multiple myeloma" (CLIOMMICS). As part of this work we reviewed decision support methods for applicability in systems medicine.

Methods

We analyzed literature on rule-based and experience-based methods for decision support systems; especially with respect to our reference disease *multiple myeloma*. Criteria for this assessment included the capability of the approach to handle:

- high dimensional data, especially genomic data
- lack of comprehensive formal *a priori* knowledge like genomic pathways about multiple myeloma
- uncertainty in source data

Based on these factors we selected a decision support method for our prototype implementation.

Results

Rule-based approaches such as rules engines and deductive classifiers are based on the availability of a formal representation of knowledge such as an ontology. Decisions are inferred by analyzing relations modelled into the knowledge base. In contrast, approaches like machine learning or case-based reasoning (CBR) are based on prior experiences made in treatment of the disease. In such an approach, no prior formal representation of knowledge is necessary.

For multiple myeloma, we currently do not have a comprehensive model of the disease available. Thus, we decided to implement the IT system for CLIOMMICS with a CBR decision support system. Publications suggest that CBR is applicable to genomic data despite their high dimensionality [1–3].

Currently, a prototype decision support system based on CBR has been developed for CLIOMMICS. A case basis of about 3000 patients with clinical and genomic data is available.

Discussion

A major challenge in this approach is the adequate definition of similarity functions reflecting the heterogeneous character of the parameter set. Especially in genomic data, significant noise is expected. In the future we may investigate a hybrid system combining CBR and a rule-based approach to add medical knowledge from sources like medical guidelines.

Conclusion

CBR seems to be a promising approach to implement decision support for systems medicine.

Acknowledgement

CLIOMMICS is funded by the German Ministry of Education and Research within the e:Med initiative, grant 01ZX1309A.

References

[1] Yao B, Li S. ANMM4CBR: a case-based reasoning method for gene expression data classification. Algorithms Mol Biol 2010; 5:14.

[2] Díaz F, Fdez-Riverola F, Corchado JM. gene-CBR: A case-based reasonig tool for cancer diagnosis using microarray data sets. Comput Intell 2006; 22(3-4):254–68.

[3] Arshadi N, Jurisica I. Data mining for case-based reasoning in high-dimensional biological domains. IEEE Trans Knowl Data Eng 2005; 17(8):1127–37.

Address for correspondence

Email: matthias.ganzinger@med.uni-heidelberg.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-952

Bridging the Gap between Clinical Practice Guidelines and Archetype-Based Electronic Health Records: A Novel Model Proposal

Diego Garcia[a], Claudia Maria C. Moro[a], Lilian Mie M. Cintho[a]

[a] *Graduate Program in Health Technology, Pontifical Catholic University of Paraná, Curitiba, Brazil*

Abstract

The lack of a unique, standardized format for representing data and knowledge is one of the existing difficulties to integrating decision support into Electronic Health Records (EHRs). Objective: Propose an archetype-based model to allow the integration of Clinical Practice Guidelines (CPG) and EHRs; design and implement this proposed model. Results: A generic model was designed for the integration of CPG into EHRs, and an archetype-based EHR for Chronic Kidney Disease Prevention based on rules from CPGs, was made as a proof of concept of this novel integration.

Keywords:

Archetypes; Clinical Practice Guideline; Electronic Health Records.

Introduction

Archetypes and SNOMED-CT terminology have been used to integrate Electronic Health Record (EHR) and Clinical Decision Support Systems (CDSS), typically to represent information [1]. Although archetype architectures have been used in CDSS development [2], until now, no work has presented or described a model integrated with an EHR. Based on this, the goal of this study is to propose a model of an archetype-based EHR that allows this integration, and is implemented using the Clinical Practice Guideline (CPG) for Chronic Kidney Disease (CKD) prevention.

Methods

To integrate CPGs into an EHR using archetypes it is necessary that the system handles archetypes, rules and EHR data. To provide low coupling between components and system, we used independent services for each one of these three types of resources. To collect data based on archetypes structure we used an EHR with a component that automatically generates forms from archetype. To provide decision support, it was necessary to incorporate an inference engine for the developed EHR.

The implementation of the proposed model was carried out as a proof of concept to demonstrate the applicability of the model.

Results

The main result of this work is the implementation of the information level of the archetype dual-level architecture of the proposed model. It consisted of seven components for handling archetypes, rules and EHR data: three repositories, three services and the forms generator. The components were implemented using Java technology and a PostgreSQL Database Management System (DBMS). Representational State Transfer (REST) architectural style was used to implement the services. Also, openEHR Java Reference Implementation (RI), provided by openEHR Foundation, and GDL libraries were used to parse archetypes and rules content. The EHR was implemented as a Java Server Faces (JSF) application containing a component to automatically generate forms to collect data based on archetypes definitions. It was necessary to implement the openEHR specification using standards required by the JSF Application. JBoss Drools Expert was used as inference engine to provide decision support capabilities.

The validation of the components was conducted using existing archetypes to validate the archetype, EHR data repository and service, and the forms generator. It used openEHR and ISO/EN 13606 archetypes. All openEHR archetypes were successfully processed and forms were generated. However, the forms generator was unable to process ISO/EN 13606 archetypes as the main class of clinical archetypes (Entry) as they are more generic in this model. To validate the execution of the rules, data based on archetypes structure were collected and the EHR detected the rules contained in the rules repository that were applicable to them. The rules were executed and the resulting data were in compliance with the CPG.

Conclusion

This study presented a model of an archetype-based EHR with CPG integration that bridges the gap between these systems. The EHR was able to provide decision support based on rules and data structured in archetypes, demonstrating the feasibility and applicability of the proposed model. Along with the proposed model, the implementation of the components is also one of the contributions of this study as there are few works reporting this step, particularly the component that automatically generates forms from archetypes, which is a problem that hinders the implementation of archetype-based EHR.

References

[1] Dias RD, Cook TW, Freire SM. Modeling healthcare authorization and claim submissions using the openEHR dual-model approach. BMC Med Inform Decis Mak [Internet]. 2011 Jan

[2] Anani N, Chen R, Prazeres Moreira T, Koch S. Retrospective checking of compliance with practice guidelines for acute stroke care: a novel experiment using openEHR's Guideline Definition Language. BMC Med Inform Decis Mak [Internet]. 2014 Jan

Address for correspondence

Diego Garcia, garcia@garciad.com and Claudia Moro, c.moro@pucpr.br.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-953

Virtual Oncological Networks – IT Support for an Evidence-based, Oncological Health Care Management

Katja Heiden[a], Monika Sinha[b], Britta Böckmann[b]

[a] *adesso AG, Dortmund, Germany*
[b] *Department of Medical Informatics, University of Applied Sciences and Arts, Dortmund, Germany*

Abstract

An interdisciplinary and intersectoral coordinated therapy management along Clinical Practice Guidelines can ensure that all patients receive adequate diagnostic, treatment, and supportive services that lead most likely to optimal outcomes. Within the research project "Virtual Oncological Networks", guideline-compliant pathways are defined and enacted within a Health Care Management Platform to support treatment planning and ongoing care of oncological diseases.

Keywords:

Clinical Practice Guidelines; Quality Assurance; Oncological Networks; Health Care Management.

Introduction

Cancer treatment is a collaborative effort and requires an effective co-ordination of all health activities and professionals[1]. The implementation of a cross-institutional health care management can enhance process and outcome indicators, particularly for breast cancer patients[2]. This can be realized by developing pathways based on existing Clinical Practice Guidelines (CPGs). As part of the research project, integrated care processes for breast, colon, and prostate carcinoma are defined and enacted within a Health Care Management Platform (HCMP).

Materials and Methods

The definition of standardized care processes based on existing CPGs is done by the help of *PathGuide*[3]. In a step-by-step approach relevant CPG recommendations are annotated, formalized and used for process modeling. The guideline-compliant pathways are represented within a Health Level (HL7) version 3 model. For enactment and visualization these processes are integrated into the HCMP to enable providers of regional oncological care coordinating all aspects of a patient's physical and psychosocial needs.

Results

Each guideline-compliant process is described by a set of medical, supportive and administrative activities. In addition, collaboration across professions are specified to ensure well-coordinated processes. For enactment, the HL7 models are parsed and visualized within the HCMP. Each process is linked to treatment objectives to determine the patient group, who can be treated by this process (e.g. recurrence-free survival or analgesia). This allows the definition of different treatment options, which can be used based on the type of cancer and the patients' health conditions to individualize therapy. For example, the treatment objective recurrence-free survival for a breast cancer treatment call for a breast-conserving therapy, mastectomy or chemotherapy, while the treatment objective analgesia requires a palliative care. The whole approach is clarified in Figure 1. Next, results are evaluated with a network of oncology experts.

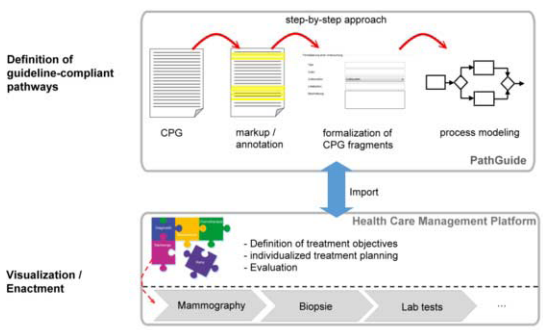


Figure 1 – Evidence-based Health Care Management

Conclusion

The use of CPGs can ensure the co-ordination of care across different professions and sectors. The definition of guideline-compliant processes allows an individualized treatment planning and can effect a continuous quality improvement of oncological care.

References

[1] Abdulrahman GO. The effect of multidisciplinary team care on cancer management. Pan Afr Med J. 2011; 9: 20.
[2] Horning SJ, Mellstedt H. Introduction: ASCO-ESMO consensus statement on quality cancer care. Ann Oncol 2006 Jul; 17(7): 1061.
[3] Böckmann B, and Heiden K. Extracting and transforming clinical guidelines into pathway models for different hospital information systems. Health Inf Sci Syst 2013, 1:13.

Address for correspondence

Prof. Dr. Britta Böckmann, University of Applied Sciences and Arts, Emil-Figge-Str. 42, Dortmund, britta.boeckmann@fh-dortmund.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-954

Computer-Interpretable Clinical Guidelines: A Review and Analysis of Evaluation Criteria for Authoring Methods

Soudabeh Khodambashi[a], Laura Slaughter[b], Øystein Nytrø[a]

[a] *Department of Computer and Information Science, Norwegian University of Science and Technology, Norway*
[b] *Oslo University Hospital, Oslo, Norway*

Abstract

There are a variety of authoring tools and methods for producing computer-interpretable clinical guideline (CIG). This work is a review of the evaluation of tools and methods currently in use to author CIGs. The aim of this paper is to present the results of a literature review on the evaluation criteria. Both controlled database search and a subsequent snowballing were used to identify relevant literature. The evaluation criteria and evaluation methods of CIG-related themes were manually identified in the found literature. Based on the 32 relevant papers found, 68 evaluation criteria were identified which were then classified into ten themes. We identified the most and least frequently mentioned areas of concern in evaluation which indicate areas that have been neglected in system evaluation.

Keywords:

Computer-interpretable clinical guidelines; Evaluation criteria.

Introduction

A multitude of tools and methods have been developed to support authoring of computer-interpretable clinical guidelines (CIGs) [1]. The tools that are designed for encoding guidelines have been evaluated in numerous studies [2]. The criteria they have chosen in their evaluation methodologies are based on these authors' desire to ensure effective implementation and to assess the impact of the tools. This paper reviews the themes, criteria found in a selection of reported research on CIG. The review covers evaluation studies as well as comparative studies. The scope of this paper is limited to research describing modelling languages and the modelling/encoding process. Papers related to application of CIGs, systems for executing CIGs or clinician's compliance with recommendations are out of the scope.

Methods

Our research method was literature review in the Journal of Biomedical Informatics, Journal of the American Medical Informatics Associations, Methods of Information in Medicine, International Journal of Medical Informatics, Artificial Intelligence in Medicine and Digital Bibliographic Library Browser. A total number of 131 papers were included based on title and abstract. In a second phase, other related papers for review were selected based on snowballing method to the identified articles in addition to the searches in the databases.

Results

From the 32 fully reviewed papers, we identified a total of 68 evaluation criteria. We identified similar evaluated criteria and grouped them in order to generate the evaluation themes covered in research studies. The evaluation themes were subdivided into 10 main categories: knowledge extraction performance, tool functionality, modelling perspective, user's effect on modelling, didactic support, usability, guideline verification and validation, integration, maintenance features and comparative analysis on encoding process.

Discussion

The themes receiving most attention based on the results and the numbers of identified criteria were tool functionality and evaluation on modelling perspective. At the other end, the themes given least attention were knowledge extraction performance, verification and validation, and maintenance features. The limited number of studies in the domain of usability indicates the least concern in the evaluation of the tools and methods in the literature. Based on the number of proposed tools and methods in the domain of CIG, we expected to find more articles on usability testing of the tools.

Conclusion

Overall the results of this review shows that the field of CIG is still mainly concerned with technical and functional aspects of the technology and has yet to move to a stronger focus on its users. A comprehensive overview of appropriate evaluation methods and criteria which consider different aspects is necessary to facilitate further work in the field. The results motivate the development of a comprehensive evaluation framework covering more aspects of tool and method properties and qualities.

Acknowledgments

I am grateful for the support of the EviCare project (http://www.evicare.no), Norwegian Research Council grant no 193022, contributing project partners and Eric Monteiro.

References

[1] de Clercq PA, Blom JA, Korsten HHM, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med. 2004;31(1):1-27.

[2] Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomed Inform. 2013;46(4):744-763.

Address for correspondence

Soudabeh@idi.ntnu.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-955

Internal domain-specific language based on Arden Syntax and FHIR

Eizen Kimura[a], Ken Ishihara[a]

[a] *Dept. Medical Informatics of Medical School of Ehime University, Japan*

Abstract

We demonstrate the concept of Arden Syntax, which is like Domain Specific Language (DSL), for describing decision support rules that take advantage of modern programming language concepts. Using object relational mapper and meta programming enables the conceptualization of FHIR resources from various data sources, and reduces the complexity of knowledge retrieval in decision rules.

Keywords:

CDSS, Arden Syntax, FHIR, DSL, ESB, DWH

Introduction

The HL7 International Arden Syntax (AS) for Medical Logic Modules (MLMs) [1] was developed to standardize the representation of medical knowledge for clinical decision support. However, the widespread adoption of AS has been a problem because of two issues. First, it is difficult to supply skilled developers because AS is an external domain-specific language (DSL), thereby limiting learning opportunities. Second, there is no standardization of data access methods in AS, and every healthcare entity has to implement specific queries for their data source. There is no unified method to manage bulk messages from a variety of data sources, such as the enterprise service bus (ESB) or data warehouse (DWH). This has become known as the curly braces problem. These issues decrease the portability of clinical decision rules. In this study, we developed a novel scheme to overcome these issues.

Methods

We used Ruby meta programming to implement internal DSL [2] that closely conformed to the AS specification. Mapping between DWH and Fast Healthcare Interoperability Resources (FHIR was developed with Ruby on Rails (RoR) to absorb various database structures (Fig. 1). HL7 messages exchanged between systems were converted into FHIR objects through ESB. To manipulate medical knowledge from DSL, we extended the FHIR object so that it was able to transparently query against DWH to retrieve historical medical records.

Results

We were able to translate the clinical decision support rule written in AS into our developed DSL with equivalence and expressivity. We avoided using custom SQL queries against the specific schema of the database, and retrieved information using a standardized concept and query method. HL7 messages, as transaction data and historical medical records, were unified into FHIR objects, so we were able to describe a rule for inquires about events from past to present, and from every data source in a unified way. This significantly reduced the amount of data query code and made simple rules.

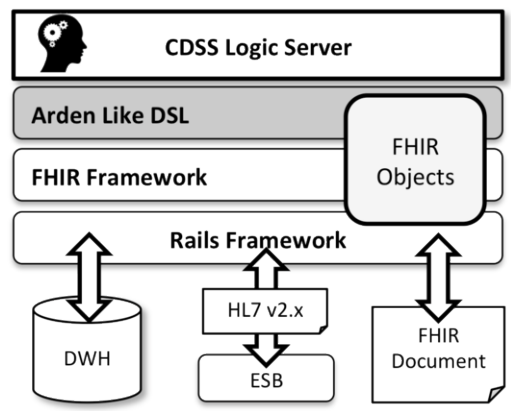


Figure 1. CDSS DSL framework with FHIR Objects

Discussion

Compared to a pure AS DSL, our approach has the advantage of availability of a modern development environment; for example, test-driven development may contribute to building solid rules representing complex clinical guidelines. And we can take advantage of the many developers, and entry-level work for implementing clinical decision rules may be reduced.

Acknowledgments

This work was supported by a Grant-in-Aid for Young Scientists (B) No. 24790504.

References

[1] Hripcsak, George, et al. “The Arden Syntax for Medical Logic Modules.” Symposium on Computer Applications in Medical Care. AMIA, 1990.

[2] Cunningham, H. Conrad. “A little language for surveys: Constructing an internal DSL in Ruby.” Proc. of the 46th Annual Southeast Regional Conference ACM, 2008.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-956

Computationally Comparing and Analyzing All Published Scoring Systems for Diagnosis of Disseminated Intravascular Coagulation

Fabrício S. P. Kury, M.D., James J. Cimino, MD

National Library of Medicine, US National Institutes of Health, Bethesda, Maryland, USA

Abstract

The clinical literature presents four different scoring systems (SS) for the diagnosis of disseminated intravascular coagulation (DIC) by four institutions: ISTH, JMHLW, JAAM and KSTH. In this study a Java program was written to retrieve medical records from the MIMIC-II database and apply the criteria of all four. The program then quantified the agreement of each DIC SS with each other and demonstrated notorious dissent. Furthermore, the average internal composition of each score was also quantified. All source code produced is available for download at https://github.com/fabkury/hedicim.

Keywords:

Disseminated intravascular coagulation; EHR databases; MIMIC-II; Clinical Decision Support Systems

Introduction and Methods

Disseminated intravascular coagulation is a **c**ondition with reported mortalities above 40% [1] for which four different organizations have published diagnostic criteria (scoring systems – SS): the International Society of Thrombosis and Hemostasis (ISTH), the Japanese Ministry of Health, Labour and Welfare (JMHLW), the Japanese Association for Acute Medicine (JAAM), and the Korean Society of Thrombosis and Hemostasis (KSTH) [1]. All patient data required by these SS are available in structured form in the MIMIC-II database, permitting a computer program to readily apply the clinical logic and establish diagnosis of DIC.

A Java program was written to retrieve, from MIMIC-II, hospital admissions of non-neonates containing an ICD-9 code of a condition known to predispose the patient to DIC [2]; read all needed patient data; then apply each DIC SS, thereby identifying the moments of start and end of diagnosis as per each DIC SS. The program calculated the periods of time and cases of agreement between the DIC SS, and, additionally, reported the average proportion that each individual criteria inside each DIC SS represented towards their final scores.

Results and Discussion

A total of 36 different data elements were found to be needed from MIMIC-II. From 8284 hospital admissions of non-neonates found eligible, 1584 cases were identified with at least one DIC SS affirming a diagnosis earlier than 12 hours before patient death. Table 1 demonstrates how much each SS agreed with each other. The conclusion from our study is a great disagreement among seemingly similar clinical diagnostic criteria. In accordance with the previous literature, the JAAM SS was found to be the most sensitive, and the ISTH SS the most specific. Moreover, only 171 (10,8%) of the hospital admissions with any DIC diagnosis had been ICD-9-coded for DIC (code 286.6). The average internal composition of the scores is presented in the poster, which also demonstrates some encountered caveats to the conversion of a paper guideline into a computer, and exposes some difficulties which we faced that are inherent to retrospective analyses of EHR databases. This study exemplifies the enabling power of structured data for clinical decision support systems, because our computerized analyses would be much less feasible if the DIC scoring systems were not mostly based on data that are commonly available in structured form in EHRs.

Acknowledgments

This work was supported by intramural research funds from the NIH Clinical Center and the National Library of Medicine.

References

[1] Nisio MD et al. Diagnosis and treatment of disseminated intravascular coagulation: Guidelines of the Italian Society for Haemostasis and Thrombosis (SISET). Thrombosis Research 129 (2012) e177–e184.

[2] Thachil J and Toha CH. Current concepts in the management of disseminated intravascular coagulation. Thrombosis Research 129, Supplement 1 (2012) S54–S59.

Address for correspondence

Fabrício S. P. Kury, M.D., U.S. National Institutes of Health
Office: +1 301-435-3183, E-mail: fabricio.kury@nih.gov

Table 1 – Agreement between DIC scoring systems (SS): Overlapping periods of time (Overlapping Cases)

Denominator \ Numerator	ISTH	JMHLW	JAAM	KSTH
ISTH	*100% (100%)*	83,3% (85,6%)	92,8% (94,3%)	61,5% (66,5%)
JMHLW	28,7% (54,4%)	*100% (100%)*	96.2% (99,0%)	34.3% (47,9%)
JAAM	8,2% (24,5%)	24.7% (40,5%)	*100% (100%)*	9.0% (22,0%)
KSTH	52,0% (74,2%)	84.4% (84,2%)	86.1% (94,5%)	*100% (100%)*

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-957

Precedent Approach to Decision Making in Clinical Processes

Malykh Vladimir L.[a], Guliev Yadulla I.[a]

[a] *The Medical Informatics Research Center, Ailamazyan Program Systems Institute of RAS, Russia*

Abstract

This poster describes the results of promising research in the field of clinical processes management and decision making support. The authors formulated common scientific problems connected with the modelling of treatment processes. The research is supported by grants from the Ministry of Education and Science of the Russian Federation (the project RFMEFI60714X0089).

Keywords:

Medical information system; Model of clinical processes; Decision-making system in medicine.

Introduction

For a long time, the attempts to use mathematical methods in medicine, in particular for solving the problems of establishing the right diagnosis, forecasting the development and the outcome of clinical processes, preventing any critical situations, making decisions and managing clinical processes in general, lead to partial results that only had practical meaning in limited contexts of a particular nosology. Multiple difficulties connected with construction of the clinical process for a particular nosology, as well as the difficulty of identification of parameters for such models, did not allow for a wide adoption of clinical process models in MIS. It may look like the model approach would prevail but there is another approach to the same problems which may come forward, namely the precedent approach, and there are sufficient grounds to take it into consideration.

Methods

A recognized and experienced Russian specialist in the field of data processing methods N.G. Zagoruyko recognizes that there are two common application-oriented approaches for building systems for decision making support based either on precedents or on models. Both approaches are considered of equal worth, and both have their advantages and shortcomings. In the model approach, “just as is the case about any generalization, some features of the system behavior are lost at each particular point of decision space. When decision is based on precedent, as our experience shows, local specifics are taken into account, which allows for more correct decisions (N.G. Zagoruyko).

Results

Malyh et al. [1] offer precedent approach for building a model of clinical processes of the class of controlled stochastic processes with memory. The model is based on two postulates: modeling in the class of stochastic processes and the precedential effect of decision-making in medicine. The conceptual basis of the model has a clear, meaningful interpretation for medical professionals.

Conclusion

As a conclusion, we present a plan of tasks to be addressed in order to implement this large project aiming at decision making support in clinical processes: (1) Development of a model that could be used as a standard theory of the knowledge system and diagnostic and treatment process. (2) Development of methodology for formalization and normalization of clinical facts ensuring fast changes of the evaluation system and terminologies, and standardization of language describing facts. (3) Development of a scaled architecture of the clinical data bank that could be used in separate medical facilities and nationwide as a national standard. (4) Development of methods for relevant search of clinical precedents. (5) Development of methods ensuring the relevance and integrity of the clinical data bank for long periods of time. (6) Development of software tools for decision making support based on the precedent approach.

References

[1] Malyh V.L. [Malykh V.L.], Guliev Ja.I. [Guliev Y.I.] Upravljaemyj stohasticheskij precedentnyj process s pamjat'ju kak matematicheskaja model' lechebno-diagnosticheskogo processa [Mathematical model of clinical processes implemented in the class of stohastic Markov control processes with memory]. Informacionnye tehnologii i vychislitel'nye sistemy [Information technologies and computing systems] 2014: 2: 60-72.

Address for correspondence

152020, Pereslavl-Zalessky, Russia, Ailamazyan Program Systems Institute of RAS. Malykh Vladimir L. Ph.D., Head of Laboratory of the Medical Informatics Research Center, e-mail: mvl@interin.ru

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-958

Semantic Interoperability in Clinical Decision Support Systems: A Systematic Review

Luis Marco-Ruiz[a,b], Johan Gustav Bellika[a,b]

[a] *Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway*
[b] *Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø*

Abstract

The interoperability of Clinical Decision Support (CDS) systems with other health information systems has become one of the main limitations to their broad adoption. Semantic interoperability must be granted in order to share CDS modules across different health information systems. Currently, numerous standards for different purposes are available to enable the interoperability of CDS systems. We performed a literature review to identify and provide an overview of the available standards that enable CDS interoperability in the areas of clinical information, decision logic, terminology, and web service interfaces.

Keywords:

Clinical Decision Support Systems; Semantic Interoperability; Terminologies; Clinical Models; Ontologies.

Introduction

The interoperability and reuse of Clinical Decision Support (CDS) systems with other health information systems has become one of the main limitations to their broad adoption[1–3]. Definition of the standards for EHR interoperability, decision logic, and terminologies has represented a step towards CDS systems' reuse. For the CDS systems to be reusable across organizational boundaries a high degree of semantic interoperability (SiOp) is needed. We aim to provide an overview of the main types of standards covered in the literature, facilitating a clearer picture of the available standards to the CDS community for building interoperable systems.

Methods

We searched PubMed, IEEE Xplore, and ScienceDirect for the SiOp approaches in CDS. . Inclusion criteria were: (a) Study describing a CDS with SiOp capabilities with other systems; (b) paper covering mechanisms for the CDS reuse by means of its information, knowledge, or logic representation.

Results

We includeed 26 papers in the review available in [4]. 23% (n=6) covered the application of medical decission logic and guidelines representation formalisms; 62% (n=16) presented the use of clinical information standards; 38% (n=10) made use of semantic web technologies such as ontologies; 42% (n=11) covered the use of standard terminologies; and 19% (n=5) proposed the use of web services for CDS encapsulation or new techniques for the discovery and provenance specification of systems.

Discussion

The results show that the most popular terminologies were – SNOMED-CT covered in 82% (n=9) of the studies reviewed; LOINC in 55% (n=6) of the studies; RxNorm in 36% (n=4); and UMLS in 27% (n=3). NDF-RT was one of the less commonly used terminologies. With regards to the information standards, we found that 37% of the studies covered HL7 RIM, 31% openEHR, 25% HL7 CDA/CCR and 12% covered other standards like Intermountain CEMs or HL7 VMR. Among the studies covering the use of semantic web technologies 30% (n=3) used the ontologies for knowledge representation; 30% (n=3) used the ontologies as integration mechanisms between different conceptual models or between the EHR and the CDS logic; and 20% (n=2) used logic specification mechanism with SWRL.

Conclusion

Among all the papers we identified five main fields of work that contributed to address SiOp of the CDS systems: information standards with HL7 RIM as the most extended, SNOMED-CT as the most covered terminology, medical logic specification formalisms with the Arden syntax the most spread, knowledge management using semantic web technologies, and web services.

References

[1] Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10:523–30. doi:10.1197/jamia.M1370.

[2] Ahmadian L, van Engen-Verheul M, Bakhshi-Raiez F, Peek N, Cornet R, de Keizer NF. The role of standardized data and terminological systems in computerized clinical decision support systems: Literature review and survey. International Journal of Medical Informatics 2011;80:81–93. doi:10.1016/j.ijmedinf.2010.11.006.

[3] Huff SM, Oniki TA, Coyle JF, Parker CG, Rocha RA. Chapter 17 - Ontologies, Vocabularies and Data Models. In: Greenes RA, editor. Clinical Decision Support (Second Edition), Oxford: Academic Press; 2014, p. 465–98.

[4] http://telemed.custompublish.com/shaman-archetype-based-modelling-of-symptom-based-decision-support-systems.5389123-247950.html (accessed March 25, 2015).

Address for correspondence

Norwegian Centre for Integrated Care and Telemedicine, University hospital of Northern Norway, P.O. Box 35, N-9038 Tromsø, Norway (e-mail: Luis.Marco.Ruiz@ telemed.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-959

Development of a Mobile System Decision-support for Medical Diagnosis of Asthma in Primary Healthcare – InteliMED

Júlio Menezes Jr.[a], Cristine Gusmão[a,b]

[a] *SABER Tecnologias Educacionais e Sociais Research Group, Universidade Federal de Pernambuco, Recife, Brazil*
[b] *Departamento de Engenharia Biomédica, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, Brazil*

Abstract

The structure of public and primary healthcare in Brazil is organized in a way to provide decentralized services. In theory, this scenario could enable the usage of mobile devices integrated with information systems of several purposes. In addition, there is a need of decision-support tools that are based on collected evidences, once the professional of primary healthcare, which essentially has general knowledge (non-specialist). Therefore there is a need of information that support the decision-making process on more specific contexts. This paper presents the proposal, experience of development and application of the InteliMED, a decision-support system to asthma diagnosis of children and adolescents through decision-trees and mobile devices (smartphones and tablets).

Keywords:
Decision-support System; Mobile Health; Case Study.

Introduction

Services of primary public healthcare in Brazil are essentially decentralized and performed by professionals that have general knowledge, once this kind of service is focused on preventive actions to the population. Considering the context of primary healthcare in Brazil, we realize that there is an organizational structure that allows the usage of mobile devices, because it works on distributed healthcare centers, in which the teams of professionals are responsible by a small part of the population and,they can perform homecare.

In this light, this work presents a proposal of medical decision-support system decentralized through the usage of mobile devices – InteliMED [1], the experience of the development of this system including a case study.

InteliMED

InteliMED is a decision-support system whose goal is to provide assistance to the primary healthcare centers for a remote support of medical diagnosis. InteliMED system aims to use decision-tree artificial intelligence technique applied to programs of communities' assistance [1].

Case Study

To evaluate the InteliMED system and its decision-tree, we performed a case study with two objectives:

1. To verify the hit rate of the generated decision-tree on the mobile application;

2.To identify points of improvement on the whole system.

InteliMED were evaluated in the ambulatory of allergy of *Hospital das Clínicas* from *Universidade Federal de Pernambuco* (HC/UFPE). The ambulatory of allergy of HC/UFPE is a specialized center of allergy and has many cases of breath allergies in children and adolescents, including Asthma. This case study was performed between June 15th and October 25th 2012 with the support of specialist doctors, in which 43 patients with ages from 5 to 18 years were consulted.

Main Results and Ongoing Work

From a total of 43 patients, the InteliMED got 36 correct and 7 wrong diagnoses (i.e., a hit rate of approximately 83,72%), which is an acceptable result. According to the opinion of the doctors, we got 25 asthmatics and 18 non-asthmatics patients.

Another interesting result is: in all the wrong diagnoses the system informed that the patient had Asthma, in which the correct result (the doctor opinion) is "No Asthma". This means that the InteliMED has been successful in all cases of non-asthmatics patients. InteliMED system brings the importance of decision-support tools in primary healthcare, concerning assistance and permanent education of healthcare professionals that work in this field.

Currently we are working on the development of educational resource for mobile devices for physical educators for older adults assessment. An initial work was developed with a group of older adults and now we are raising the system requirements.

References

[1] Menezes Jr. JV, Gusmão C. InteliMED Proposta de Sistema de Apoio ao Diagnóstico Médico para Dispositivos Móveis. iSys: Revista Brasileira de Sistemas de Informação, v. 6, 2013; p. 44-61.

Address for correspondence

Júlio Menezes Jr. (juliovenancio@gmail.com), Espaço SABER, sala 04, Centro de Convenções, Universidade Federal de Pernambuco – UFPE. Av. dos Reitores - Cidade Universitária, Recife - PE, Brazil. Zip code: 50741-000.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-960

Using Discrete Event Simulation to predict KPI´s at a Projected Emergency Room

Pablo Concha[a], Liliana Neriz[b], Danilo Parada[a], Francisco Ramis[a]

[a] *Centro Avanzado de Simulación de Procesos; Universidad del Bío-Bío, Chile.*
[b] *Departamento Control de Gestión y SI, Universidad de Chile, Chile.*

Abstract

Discrete Event Simulation (DES) is a powerful factor in the design of clinical facilities. DES enables facilities to be built or adapted to achieve the expected Key Performance Indicators (KPI´s) such as average waiting times according to acuity, average stay times and others. Our computational model was built and validated using expert judgment and supporting statistical data. One scenario studied resulted in a 50% decrease in the average cycle time of patients compared to the original model, mainly by modifying the patient´s attention model.

Keywords:

Emergency Department Design, Healthcare Simulation, KPI..

Introduction

The performance of an Emergency Department (ED) is affected by the integration of its physical design, staffing, and the equipment and service model. Inadequate design of any of these factors may result in poor future performance of the facility. Because of the complexity among the different processes and the variability present in EDs, this situation can be studied by using a DES model.

Our study was based on the use of DES tools, which are used to study KPI´s such as average cycle time, time to first medical evaluation, and utilization factor of the different rooms and resources. The model was developed using the simulation software *Flexsim Healthcare*. The methodology of the study followed seven steps: Problem formulation, Construction and validation of the conceptual model, Model construction, Verification and validation of the computer model, Design of experiments, Analysis of results and finally Documentation and presentation of results.

The flow of patients through the ED, also called carepathways, depends on the type of patient and activities which may include: admission, triage, medical assessment, treatments/procedures, diagnostic support and discharge from the ED.

A total of fourteen carepathways for the patients were defined.

1. EKG, hydration, sampling, observation
2. Nebulization
3. Intramuscular treatment, hydration, sampling
4. EKG, observation
5. hydration and sampling
6. Imaging, cast
7. Hydration, minor surgery, imaging
8. Hydration, minor surgery
9. Minor surgery, nurse procedure
10. Imaging, nurse procedure
11. Breath analysis (Example in Figure 1)
12. Medical procedure, wound treatment
13. Sampling, imaging
14. Nurse procedure

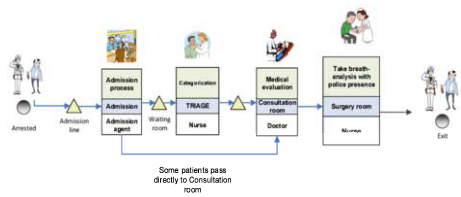


Figure 1– Sample carepathway for a breath analysis

To estimate the demand to be used in the model, the data was extracted from the database of a nearby hospital, located in the metropolitan region of Chile.

Results

The original average cycle time of patients was approximately 15 hours. The examination room was the utilization factor with the highest percentage, due to the fact all carepathways analyzed passed through this room. Additional places that showed a high utilization factor were the procedure and observation rooms, which in practice are used to perform similar activities as the examination room. Additionally, eleven of the fourteen patient carepathways passed through these rooms.

Conclusions

The simulation study conducted to analyze the future ED aided the decision making of the hospital management by allowing them to generate scenarios and analyze the possible results before implementing them. These scenarios were defined by changing the amount of stretchers in two areas and changing the attention model. Afterwards these scenarios were evaluated according to the average cycle time of the patients. The optimal reduction of cycle time of patients was the scenario where a stretcher was added in the procedure room and the attention model was changed. This decreased the cycle time from 14.8 hours to 3.7 hours in scenario 1 and the time to doctor from 11.5 hours to 1.1 hours.

References are available on request

Address for correspondence

Francisco Ramis Lanyon; E-mail: framis@ubiobio.cl; Phone: +56-41-3111755

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-961

Clinical Decision Support Using Electronic Medical Records: For the Improvement of Diabetes Care and Proper Use of Insulin for Inpatients

Ryoma Seto[a], Susumu Wakabayashi[b]

[a] *Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare Informatics, Tokyo, Japan*
[b] *Division of Drug Information, Department of Pharmacy, Kyorin University Hospital, Tokyo, Japan*

Abstract

The aim of the study is to develop a scheme of a decision support system concerning insulin intervention for inpatients. Transaction data for 32,637 inpatients were collected from the EMR. As a result, antidiabetic agents were not taken by 38.9%–41.7% of patients with a Disease Complicated by DM. It is recommended that the EMR should provide a suggestion about insulin level for diseases with DM as a complicating factor.

Keywords:

Electronic Health Records, Drug Utilization Review, Diabetes Complications

Introduction

Diabetes prevalence is increasing in Japan, from 7.4 million in 2002 to 9.5 million in 2012. All aged ≥40 are required to have a medical checkup that includes a diabetes test. However, some patients ignore this and their condition deteriorates. The aim of this study is to develop a framework of medical information and a decision support system concerning insulin intervention for inpatients based on electronic medical records (EMR), for the purpose of preventing diabetes (DM) patients from progressing towards hemotoxic dialysis (HD).

Methods

The development was based on 3 major hospitals around Tokyo. Transaction data for 32,637 inpatients between 2009 and 2011 with regard to prescription and medical procedures related to DM were collected from EMR. Based on the data, the drug use recommendations were assessed.

Results

1) Prevalence of DM: Among the inpatients, 2.0%–3.3% had their disease registered as DM. Of all patients with DM, 21.8%–78.0% had DM, 12.6%–54.1% had a disease complicated with DM (e.g. CKD) and 9.3%–26.1% had other diseases (e.g. cancer) associated with DM.

2) Usage of Insulin and Oral Hypoglycemic Agents for all Patients with DM: Of patients with DM, 67.3%–84.3% were using insulin and/or oral hypoglycemic agents, 22.7%–28.8% of them were using only insulin, 11.0%–46.2% of them were using only insulin, and 19.2%–27.3% of them were using both insulin and oral hypoglycemic agents.

3) Drug Use of Diabetic Patients: Among the patients, 21.4%–38.9% were using only insulin, 22.7%–35.4% of them were using oral hypoglycemic agents and 32.8%–33.3% of them were using both. No antidiabetic agents were taken by only 5.6%–9.9% of patients.

4) Drug Use of Patients with a Disease Complicated by DM: 33.3% of patients were using only insulin, 16.6% of them were using oral hypoglycemic agents and 8.3%–11.1% of them were using both. Antidiabetic agents were not taken by 38.9%–41.7% of patients and 49.6% of them were not implementing self-measurement of blood glucose (SMBG).

5) Drug Use of other Disease Patients: Of patients registered as having DM, 11.5%–22.7% of them were using only insulin, 17.6%–25.0% of them were using oral hypoglycemic agents and 6.3%–47.1% of them were using both. No antidiabetic agents were being used by 43.8%–76.5% and 67.1% of them were not implementing SMBG.

6) Possibility of Uncertain Disease Registration: Of the patients implementing SMBG, those patients having DM or complications related to DM was 10.6%–29.0%.

Discussion

1) Proper Use on Antidiabetic Agent for Patients with DM: Most patients with DM were given oral hypoglycemic agents and/or insulin. The situation is a natural one because the physicians in charge are mostly specialized endocrinologists.

2) Proper Use on Antidiabetic Agent for Patients with a Disease Complicated by DM: Of patients with a disease complicated by DM, the percentage using insulin was lower than in patients with DM. However, the pathological conditions of patients with a DM complicated disease are more serious than in patients with only DM.

3) Proper Use of Other Disease Patients: With regard to patients with other diseases, some patients ignored the use of oral hypoglycemic agents, insulin and SMBG. This situation may arise from the physicians' specialty.

Conclusion

As a result of this study, we suggest that intervention via EMR is necessary to provide recommendations of insulin level for diseases with DM as a complicating factor, and any antidiabetic agents and SMBG for patients with other diseases. Information may be provided in two ways. The first is the caution in EMR for suggesting the proper use. The second is the distribution of information to physicians, pharmacists and medical officer assistants to use antidiabetic agents with the aim that more inpatients may be given antidiabetic agents. Based on the 2012 national health and nutrition examination, 26.8%–30.4% of people with suspected DM did not have DM care.

This work was supported by JSPS KAKENHI (No.24790628).

Address for correspondence

Ryoma Seto (E-mail: r-seto@thcu.ac.jp)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-962

Why should I? – Acceptance of Health Information Technology Among health professionals

Joseph M.J. D'Souza[a], Inga Hunter[b]

[a] *Intensive Care Unit, Christchurch Hospital, Christchurch New Zealand*
[b] *School of Business, Massey university, Palmerston North, New Zealand*

Abstract

Applying the Technology Acceptance Model, the end user intentions to use technology applications is studied. The study finds the end users negative perception of the usefulness of the application as a major factor in its suboptimal utilisation.

Keywords:

Technology Acceptance; Health information Technology;

Introduction

Health Information Technology (HIT) and its applications have been promoted as a solution to the problems of inefficiencies, medical errors, and rising costs in delivery of health care [1]. However the benefits of the HIT on health outcomes are a product of its actual and efficient use by the clinicians [2].This research enquires into the factors influencing the clinician's intentions in the utilisation of the HIT applications. Applying Technology Acceptance Model (TAM), the extensively used model in technology acceptance studies [3], the use of the Central Line Associated Bacteraemia (CLAB) application is studied.

Methods

This study is designed as a cross-sectional quantitative study, using paper based self-completed survey questionnaire. The questionnaire is based on the instrument used by Brown, Massey, Montoya-Weiss and Burkman [4], and validated by Davis 1989 [5]. Accordingly, four variables, namely, PU, PEOU, Attitude, and intention to use the application were measured. Responses were graded on 5point Likert scale, 1 being extremely negative and 5being extremely positive. The data gathered was analysed using descriptive analyses.

Results

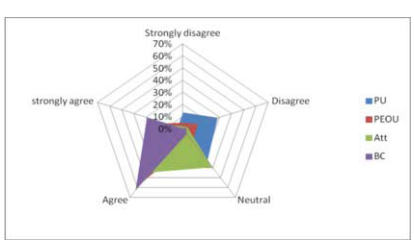


Fig.1. Average responses for all variables

Discussion

The overall findings and trends of all average responses are plotted into one graph (Fig. 1). This graph displays that the PU is the only striking outlier of all the variables with most negative responses along with a mostly neutral perception of the attitude of the users towards the application. The negative perception of the usefulness of the CLAB application and a neutral attitude towards the use of the application. The empirical evidence shows that the PU and the attitude towards use as the main factors influencing the actual use of technology [3], [5].

Conclusion

Understanding the key drivers of user acceptance of the technology and its optimal use is crucial in successful implementation of information technology, including HIT. As with other HIT applications, in order to achieve optimal use of the CLAB application it is essential to foster a positive perception of the usefulness of the application through effectively disseminating the information of its utility in achieving safer and better delivery of patient care.

References

[1] Chaudhry B, et al. (2006) Systematic review: Impact of Health Information Technology on Quality, efficiency and costs of medical care. Annals of Internal Medicine. 2006;144(10):742-52. PubMed PMID: 16702590

[2] Goldzweig CL, et al. (2009). Costs and Benefits from health information technology: New trends in literature. Health Affairs. 2009;28(2):w282-w289. doi: 10.1377/hlthaff.28.2w282

[3] Mathieson K, Peacock E, Chin WW. (2001). Extending the technology acceptance model: The influence of perceived user resources. SIGMIS database. 2001;32(3);86-112.

[4] Brown SA, Massey AP, Montoya-Weiss MM, Burkman JR. (2002). Do I really have to? User acceptance of mandated technology. European Journal of Information Systems. 2002;11(4), 283-295.

[5] Davis FD. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. MIS Quarterly. 1989;13(3):258-68. doi: 10.2307/2409008.

Address for correspondence

Joseph D'Souza, Intensive Care Unit, Christchurch Hospital, Private Bag No. 4710, Christchurch-8140, New Zeland Ph.: 0212044922; Email: joseph.d'souza@cdhb.health.nz

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-963

Diffusion of innovation: Telehealth for care at home

Sharon Levy

Nursing Studies, University of Edinburgh, Scotland, UK

Abstract

The 'care at home' study focused on a Scottish telehealth service, which was designed to support children with palliative and complex care needs. Using the diffusion of innovation theory, this poster highlights the differences between the way telehealth is used in the public sector and in a third sector or a voluntary organization. Analysis of the data, taken from interviews with key stakeholders, illuminate barriers and solutions as noted by clinicians who see the clear benefits and potential risks of telehealth use at home. In conclusion, it is argued that a strategic steer towards a culture of innovation is needed to support effective use of telehealth in clinical practice. Senior managers in the National Health Service in the United Kingdom need to 'unleash' the goodwill of staff who are eager to exploit innovation in clinical practice.

Keywords:
Telehealth; Palliative care; Complex needs; Peadiatrics.

Introduction.

The stated aims of the study were to trial telehealth technology, setup the infrastructure and associated clinical pathways, and study the care outcomes of the children with complex and palliative care needs. We intended to explore remote care in a context where often 'high touch' rather than 'high tech' is most appropriate. The palliative care arm of the study was conducted with Paediatric Oncology Outreach teams, working in the National Health Service (NHS). The phase concerning children and young people with complex needs was supported by a third sector organisation, the Scottish Spina Bifida Association (SSBA). Whilst elements from each section of the study have been reported elsewhere (1) (2), this poster compares the outcomes of each segment and identifies potential barriers and enablers to the adoption of telehealth. The cultural differences between the organisations involved, and the impact they have on the diffusion of innovation, are considered in this study.

Methods

Clinicians across four specialist care locations were involved in data collection for this study. These included oncology healthcare teams in three NHS paediatric hospitals and a nursing team at the SSBA. At the end of the study the clinical teams were asked to reflect on their overall experience, lessons learned, and their perceptions on the way the service could be further developed. This final engagment was conducted as a face to face semi structured interview which lasted up to one hour. Interviewees included seven peadiatric outreach oncology nurses, four medical consultants, as well as two specialist nurses and an outreach worker from the SSBA. Interviews were recorded, transcribed, and analyzed for common themes, and the findings are reported here.

Results

Relative Advantage: Telehealth was introduced as a tool to enhance current services, rather than replacing them. In the NHS the potential advantages were also seen as a possible threat to future roles for staff. By contrast, the staff at the SSBA perceived the new remote service as an opportunity to affirm their key role within the organisation and cement the commitment of the organisation to continue to offer specialist nursing support to users. **Complexity**: In the NHS there is a low tolerance to the failure of clinical technology. Many interviewees noted their frustration with rigid clinical IT systems and the level of support they recived from their dedicated IT helpdesk. **Trialability**: Staff in the NHS were limited to the approved and avilable set-up, within their clinical areas, which often was a large Video Conferencing (VC) suite. SSBA staff trialed a number of technologies and opted to use SKYPE which they found easy to use. This meant that many more users were able to download the software on their own machines at home.

Conclusion

A number of attributes, aligned with the diffusion of innovation theory, served to illustrate the findings from data gathered in both the public and the third sector organizations. The plan of introducing telehealth in the complex organizations, such as the NHS in the UK, would require development of a calture of innovation. It needs a shift in organizational mindset, and a significant change in the perceptions of risks and attitudes towards failure.

References

[1].Levy S, Henderson L, McAlpine C. Growing up with confidence: using telehealth to support continence self-care deficits amongst young people with complex needs. Informatics in Primary Care. 2014;21(3):113-7.

[2].Levy S, Steele G. End-of-life and living technologies, the role of telemedicine in palliative care of children and young people. PervasiveHealth; 2011.

Address for correspondence

Nursing Studies, Teviot place, Edinburgh, Scotland.
Sharon.levy@ed.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-964

Visualizing Sensor Data through an Open Platform for Connected Devices

Christian Bock[a,d], Thai Le[a], Arjmand Samuel[b], Danny Huang[b], Hilaire J. Thompson[c], George Demiris[a,c]

[a] *Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, USA*
[b] *Microsoft Research, Seattle, WA, USA*
[c] *Biobehavioral Nursing and Health Systems, School of Nursing, University of Washington, Seattle, WA, USA*
[d] *Ruprecht-Karls-University of Heidelberg, Heidelberg, Germany*

Abstract

Integrating sensor systems within the home can be a tedious process due to the challenges of deploying systems across multiple environments, customizing applications, and connecting across various devices. We demonstrate the deployment of sensors using the Lab of Things platform within a residence over a 3 month period. We developed a real time visualization application from the generated sensor data, and evaluated it through a survey with 19 older adults and caregivers. Findings can be a valuable guide to scale a study across multiple settings, and to create personalized interfaces.

Keywords:
Data Display; Remote Sensing Technology; Aged

Introduction

In the field of gerontology, the use of smart home sensor systems is growing primarily due to the appeal of unobtrusively monitoring older adult health and wellness. However, the deployment of sensor systems has often been limited to academic or laboratory settings. This is due to the challenges of managing multiple simultaneous installations, integrating different types of sensors, and disseminating meaningful information back to consumers. We describe the deployment of sensors within a multi-person residence using the Lab of Things platform for connected devices [1], a software platform that reduces the burden of deploying sensor systems across different environments while also serving as a central hub to manage connected devices, store data and analyze data. We developed a prototype using data collected from a smart home and integrated it into a real time visualization application.

Materials and Methods

We deployed the Lab of Things platform in a house of 11 residents as a demonstration of feasibility over the course of 3 months. The system consisted of one multi-sensor and two door/window sensors. These sensors included two Aeon Labs Z-wave Door/Window sensors and an Aeon Labs Multi-sensor that collected data on motion, temperature, luminosity, and humidity. We applied a Rapid Iterative Testing and Evaluation (RITE) approach in designing the visualizations. This involved two rounds of iterative prototyping with gerontological experts. We presented early mockups to a group of 10 gerontology experts to gather feedback and design guidelines. We then conducted a third evaluation through a survey with older adult- and caregiver stakeholders to inform further revisions to the designs. We developed the application through a combination of HTML, CSS, and JavaScript for the client side. We used a service-oriented architecture to communicate with the server machine, represented by the HomeHub on the Lab of Things platform. We evaluated the resulting application through an online survey with family members/caregivers and older adults. The survey consisted of static displays of the visualization with questions asking to rate the importance of different components of the display; such as type of data presented, ability to switch amongst different views by spatial location, and privacy concerns.

Results

Our initial mockups consisted of a bar chart representing activity level over time. A “Norm Activity Index” was provided as a reference for average activity level. We used color as an encoding for activity level relative to the norm in the bar chart. Nineteen participants completed the online survey. Family members/caregivers felt comfortable using the application indicating that they could quickly learn to use the application and would not need the support of a technical person. Older adults, however, indicated it would be difficult to learn to use the application, and had trouble identifying its utility. A key for older adults provided information on how the data collected could be utilized by their family members, physicians, or caregivers.

Conclusion

The evaluation of our application with stakeholders highlighted significant differences between how family members/ caregivers perceived of the application compared to older adults. Though family members found the application easy to learn and identified value in having the information available to monitor older adult wellness, older adults did not share this same view. Instead, they found the visualization complex and difficult to learn; partially confounded by an unclear sense of how the information could be used. This lack of perceived usability can be a significant barrier to engagement.

References

[1] Brush AJB, Jung J, Mahajan R, Scott J. HomeLab: Shared Infrastructure for Home Technology Field Studies. Proceedings of the 2012 ACM Conference on Ubiquitous Computing. New York, NY, USA: ACM; 2012; p. 1108–13.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-965

Automated Evaluation of Medical Software Usage: Algorithm and Statistical Analyses

Ming Cao[a], Yong Chen[a], Min Zhu[b], Jiajie Zhang[b]

[a] *School of Public Health,* [b] *School of Biomedical Informatics*
The University of Texas Health Science Center at Houston, TX, USA

Abstract

Evaluating the correctness of medical software usage is critically important in healthcare system management. Turf[1] is a software that can effectively collect interactions between user and computer. In this paper, we propose an algorithm to compare the recorded human-computer interaction events with a predefined path. Based on the pass/fail results, statistical analysis methods are proposed for two applications: to identify training effects and to compare products of the same functionality.

Keywords:

Human-Computer Interaction, TURF, Usability

Introduction

National Institute of Standards and Technology has published guidance to improve the usability of Electronic Health Records (EHR)[2], but practical software tools to archive this goal are still in the preliminary stage. Our work here was intended to provide practitioners a module of functions within TURF (task, user, representation and function), a software aiming to measure usability objectively. The current version of TURF can record user interaction such as mouse clicks and keyboard typing. The complexity of the medical applications, including EHR, usually demands a series of tasks to be completed in a pre-specified way. We defined a *path* as a sequence of human-computer interaction steps taking place in order while each step can contain possibly unordered events. An automated algorithm comparing the recorded events with a predefined standard or alternative path was needed. It saves the burden for human to watch the operation process and decide whether a user completes a task successfully or not. To analyze the results, we devised appropriate statistical methods.

Materials and Methods

Raw data were processed as following: keyboard strokes were grouped into strings, mouse clicks were associated with a widget (or window/module), and all events were indexed by their event types, element contents and attributes. Then an experiment runner could define a standard path in the following way: (1) Put several tokens into group in which order may or may not matter; (2) Insert, remove or adjust the order of steps/groups; and (3) Specify mandatory steps.We described the algorithm to compare the recorded path from a user with the standard path. To ensure the robustness of the algorithm, we dichotomized the steps into "mandatory" and "non-mandatory". The events within the mandatory steps have to take place in order and the events within the non-mandatory steps can take place without the requirement on ordering. For the non-mandatory steps, some missing events can be tolerated. Formally, a user failed if the order of steps did not match the standard path, or any mandatory step was missed. Consider two application senarios: the first could quantify how much a training session improved the average rate of correctly operating the software. To make more accurate inference, bootstrap [3] is used to estimate the variance of the log odds ratio estimator. The second scenario is to compare two EHRs that serve the same purpose but operate on two different platforms. A typical setting is one in which groups of users are randomly assigned to product A product B and then the Generalized Linear Model is applied [4]. We adjusted for other covariates using the collected demographic information.

Results

We converted system events data into a readable series of steps. A binary indicator ("pass" or "fail") to the end user was produced for the task. For users who failed the test, we highlighted the problematic area for their future improvement as well as the percentage of completing the task. Finally, an estimate of training effects or the difference of products could be given, as well as the uncertainty and statistical significance.

Conclusion

The automated evaluation algorithm we proposed makes large scale usability tests accessible to TURF users. Our in house statistical functions can quantify the training effects and product differences. The contribution we wish to make is offering the usability improvement community a ready-to-use software, rather than developing a new theory.

Reference

[1] Zhang JJ, and Walji M. TURF: Toward a unified framework of EHR usability. J Biomed Inform 2011 Dec; 44 (6): 1056-67.

[2] Schumacher, Robert M., and Svetlana Z. Lowry. "NIST guide to the processes approach for improving the usability of electronic health records." National Institute of Standards and Technology (2010).

[3] Efron B. Bootstrap Methods: Another Look at the Jacknife. Ann. Statist 1979; Volumn 7, Number 1, 1-26.

[4] McCullagh P, and Nelder J. Generalized Linear Models. 2nd ed. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 1989.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-966

Study of Screen Design Principles for Visualizing Medical Records

Kenichiro Fujita[ab], Tadamasa Takemura[a], Tomohiro Kuroda[b]

[a] *Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan*
[b] *Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan*

Abstract

To improve UX of EMR/EHR, the screen design principles for the visualization are required. Through the study of common attributes of medical records, we present four principles and show three screen designs by applying them.

Keywords:
Information visualization, UX, electronic medical records.

Introduction

Improving User Experience (UX) of EMR/EHR is important for the performance and safety of the medicine. The visualization of medical records is a pivotal problem for the implementation [1], and it should be improved by the establishment of principles based on information visualization. We discuss the screen design principles for visualizing medical records and their practical use.

Methods

Information visualization is based on the common attributes of data [2][3][4]. What are them in medical records? Each medical record has the patient whom it is about. If it loses its patient attribute, it is not clinically useful. Time is fundamental for medical care, and all medical records have their own time attribute. Types of medical records constitute the hierarchical structure in toto. First hierarchy of types is most efficient for categorizing all records. We name it primary type which is a comon attribute of medical records.

How useful these common attributes are for the screen design? To avoid confusing patients, multiple patients should not display simultaneously. In information visualization, medical records are categorized into temporal data which are frequently visualized by time-series as it is usually most efficient [2][3]. The hierarchical structure is visualized by the tree which needs two dimensions for visualization and is difficult to coexist with time-series on screen. It is not suitable that medical records are visualized by the whole types of them. But primary type is not hierarchical and can combine with time-series.

Also, information visualization can be achieved by transforming and abstracting data [4]. Thus detailed medical data should be shown by other views such as popups.

Results

From the above discussion, we propose following screen design principles for visualizing medical records. First, one view should contain single patient's data. Second, data should be deployed in time-series. Third, data should be categorized by primary type. Fourth, data should be abstracted for overview and details should be given on-demand.

By applying these principles, at least three screen designs are plausible (see Figure1). Design1: Data are arranged in the time-series and labeled by primary type. Design2: Data are arranged in matrix that has two axis: time-series and primary type. Design3: Data are arranged in the time-series, and selecting one of the points of time-series, data at the selected point of time are shown as popup which is categorized by primary type.

Figure1 – Three screen designs

Conclusion

The discussion clearly tells that medical records have the common structure for visualization. This fact gives us expectation for the universal UX of EMR/EHR. To prove the assumption, it is necessary to apply above mentioned designs to existing EMR/EHR and evaluate the feasibility of the information visualizations.

References

[1] Chittaro L. Information visualization and its application to medicine, Artificial intelligence in medicine, Elsevier, vol.22, 2001; pp.81-88.

[2] Shneiderman B. The eyes have it: A task by data type taxonomy of information visualizations, Proc. IEEE Symposium on Visual Languages '96, 1996; pp.336-343.

[3] Tufte E. The Visual Display of Quantitative Information. Cheshire: Graphics Press, 1983.

[4] Chi EH. A Taxonomy of Visualization Techniques Using the Data State Reference Model, Proc. IEEE Symposium on Information Visualization 2000, 2000; pp.69-75.

Address for correspondence

Kenichiro Fujita, 54 Kawaharacho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan. Email: kfujita@kuhp.kyoto-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-967

User-centered design to develop clinical applications. Literature review

Daniel Luna[a,b], María Quispe[a], Zulma Gonzalez[a], Alfredo Alemrares[a,c], Marcelo Risk[a], Mauro Garcia Aurelio[a], Carlos Otero[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*
[b] *ITBA, Instituto Tecnológico de Buenos Aires*
[c] *DMICE, Oregon Health & Science University, Portland, Oregon, USA*

Abstract

User-centered design is mentioned by Norman as "the need for a design that uses the natural properties of the individuals, exploiting the relationships and constraints and focusing on the needs and interests of the user, in order to make the final products usable and understandable". This is also important in health developments. The objective of this paper is to search and analyze articles in the healthcare field where user-centered design principles have been applied. We describe findings in this topic from articles published between January 1995 and September 2014.

Keywords:

User center design, electronic health record, Health information systems, Usability.

Introduction

The concept of user-centered design (UCD) has become popular in recent years due to the importance and impact that its application has had on creating more usable products.

UCD is about the need for a design that uses the natural properties of the individuals, exploiting the relationships and constraints, and focusing on the needs and interests of the user, in order to make the final products usable and understandable. The difficulty and complexity in the design of software tools for the health field has generated the need for specially trained teams to handle the particular domain. The objective of this paper is to search and analyze articles in the healthcare field where user-centered design principles have been applied.

Materials and Methods

Pubmed, Lilacs, and ACM Digital Library were used. **Inclusion criteria:** Publications that reviewed the domain related to health applications and that referred to pieces of software. Also, if human, social, and behavioral principles were taken into account in designing the user interface. **Exclusion criteria:** Domain not related to health care, development of hardware that does not include software development, and no description in the article of the process and the techniques used. **Search strategy:** We decided to use truncation strategies. The search – whenever possible – was performed using the term "user center* + design". In instances where it was not possible to use truncation strategies, we performed the search using key phrases ("User center design" / "user centered design"). Search was conducted for publications in PubMed, Lilacs, and ACM Digital Library. We established a time frame for articles published between January 1995 and September 2014.

Results

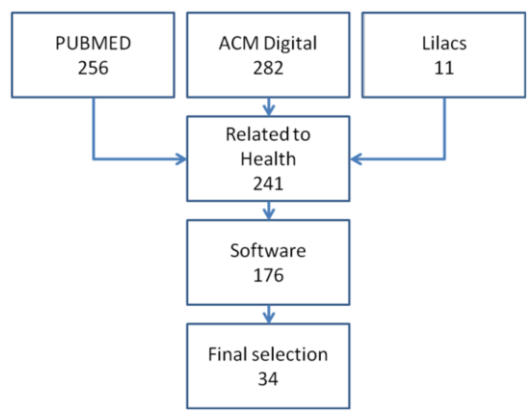


Figure 1: Flowchart with items founded.

For more detailed in the artiles analyzed see Annex 1.

Conclusion

UCD can increase the adoption and efficient use of IT tools in the field of health, reducing support, development, and maintenance cost, and ultimately increase user satisfaction and patient safety.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-968

A Model for Usability Evaluation for the Development and Implementation of Consumer eHealth Interventions

David Parry[a], Philip Carter[a], Jane Koziol-McLain[b], Jacqueline Feather[b]

[a] *School of Computer and Mathematical Science, Auckland University of Technology, Auckland, New Zealand*
[b] *Faculty of Health & Environmental Sciences, Auckland University of Technology, Auckland, New Zealand*

Abstract

Consumer eHealth products are often used by people in their own homes or other settings without dedicated clinical supervision, and often with minimal training and limited support – much as eCommerce and eGovernment applications are currently deployed. Internet based self-care systems have been advocated for over a decade as a way to reduce costs and allow more convenient care, and – because of the expectation that they will be used to reduced health cost –, by increasing self-care and avoiding hospitalization.

However, the history of consumer eHealth interventions is mixed, with many unsuccessful implementations. Many consumer eHealth products will form part of a broader complex intervention, with many possible benefits and effects on both individuals and society.

This poster describes a model of consumer eHealth assessment based on multiple methods of usability evaluation at different stages in the design and fielding of eHealth systems. We argue that different methods of usability evaluation are able to give valuable insights into the likely effects of an intervention in a way that is congruent with software development processes.

Keywords:
Consumer eHealth; Usability; Evaluation

Introduction

This paper argues that usability evaluation – in particular, observation based approaches that include investigation of the goals and characteristics of the users, and their responses to the intervention, along with the socio-technical aspects of the application – can form part of a structured trial and development approach. Frameworks for complex interventions to improve health have been defined previously. Complex clinical interventions may include software artefacts along with system redesign and behavioral change elements.

Multiple methods of usability evaluation have been shown to be effective in detecting usability issues that arise in eHealth applications. Moreover, usability evaluation before, during, and after implementation of eHealth systems has been shown to be important and useful. The usability of so-called "consumer" health products is particularly important. Consumer products are often used by people in their own homes or other settings without dedicated clinical supervision.

The use of phases in usability evaluation, just as with pharmaceutical development, emphasises the use of low-cost evaluation methods at early stages, often reducing the number of candidate interventions drastically.

Materials and Methods

TABLE 1**:** SEQUENCING OF ACTIVITY

Stage	**Domains**		
	Pharma [19]	*Complex Intervention[1]*	*Usability*
Pre-clinical	Animal testing	Theory	Requirement gathering
Phase I	Toxicity testing	Modelling	Pre-artefact testing.
Phase II	Efficacy testing	Exploratory	In vitro
Phase III	Comparison with existing treatment	Definitive Randomized Controlled trial	In vivo Experiments
Phase IV	Post- Licensing -	Long term Implementation	Benefits Redesign

A phased evaluation model should allow the early abandonment of unsuccessful designs, and the refinement of early designs before they enter the next phase. Table 1 demonstrates the analogous phases of pharmaceutical trials, complex interventions, and usability evaluation. Table 2 illustrates the corresponding usability activities during these phases.

TABLE 2

Stage	**Usability activity**		
	Evaluation techniques	*Expected outcome*	*Usability task*
Preclinical	Interviews, observation	Types of users, Goals, tasks	Requirement gathering
Phase I	Lo-fi prototyping, Heuristics	Interface design	Pre-artefact testing.
Phase II	Usability experiments	Task completion	In vitro
Phase III	Experiments and ethnographic approaches	Practical issues	In vivo Experiments
Phase IV	Surveys, feedback	Usefulness, effectiveness	Benefits, Redesign

Results

We are currently evaluating these approaches in the context of internet- and mobile-based approaches for reduction of intimate partner violence.

Address for correspondence

A/Prof David Parry, School of Computer and Mathematical Sciences, AUT, New Zealand.dave.parry@aut.ac.nz.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-969

Usability Analysis Of A Customized Documentation System For Nurse Population-Health Managers

Lincoln Sheets MD[a], Zhijian Luan[a], Tina Dillon[b]

[a] *PhD candidate, University of Missouri Informatics Institute, Columbia, MO, USA*
[b] *MEd candidate, University of Missouri School of Information Science & Learning Technologies, Columbia, MO, USA*

Abstract

The University of Missouri's population health management pilot employs 22 Nurse Care Managers (NCM) to manage medical casework for approximately 10,000 patients. We studied the NCMs' documentation system to identify interruptions to their clinical workflow, and identify missing functionality. We used an interview script and Morae software to observe and record five NCMs at work, measuring time on task, click counts, and task completion. We also documented quantitative and qualitative responses to a directed interview. All tasks were completed satisfactorily, with completion times of 443±275 seconds and click counts of 58±23. Surveys on a scale of 1 (worst) to 5 (best) scored "ease of use" at 3.4±1.1, "integration of functions" at 3.0±1.2, and "ease of learning" at 2.6±1.5. Overall, the system is functional and reliable but could be improved to support workflow.

Keywords:

Nursing Informatics; Computer Systems Evaluation; Managed Care.

Introduction

The University of Missouri's population health management pilot employs 22 Nurse Care Managers (NCM) to manage medical casework for approximately 10,000 patients. Halfway through the three-year pilot project, questions had begun to arise about the usability of the NCM documentation system. We conducted formal usability testing in order to identify specific workflow impediments and identify missing functionality.

Methods

We observed six Nurse Care Managers at their work and used Morae usability testing software to record their keystrokes, mouse movements, and mouse clicks. We also used an interview script to gather qualitative results, and five of the volunteers completed a three-item questionnaire after the test.

Results

Time on task and clicks needed

Statistic	Time (sec)	Clicks
Minimum	180	25
Maximum	920	108
Mean	443	58
Standard Deviation	275	23

Survey results (from 1=worst to 5=best)

Statistic	Q1 (ease of use)	Q2 (integration of functions)	Q3 (ease of learning)
Minimum	2	2	1
Maximum	5	5	5
Mean	3.4	3.0	2.6
Std. Dev.	1.1	1.2	1.5

Conclusion

Overall, the LIGHT^2 NCM Documentation system was functional and reliable, but we found these specific problems:

- Organization: the screen order often didn't reflect the variation in clinical encounters.
- Functionality: needed functions were present, but users had to exit the system to view related records.
- Usefulness: the users didn't find the system to be clinically useful, but saw this is a consequence of its purpose rather than a design flaw.

We can also make these general recommendations, which may be generalizable to other population-management system documentation systems:

1. Consult system users on content organization, size of fields, and order of screens.
2. Allow flexibility in re-ordering screens when appropriate.
3. Link the nursing documentation system to the related electronic medical record without exiting and re-entering.
4. Enable notes to be labeled, dated, and searched.
5. Improve performance to decrease the initial load time.
6. Add auto-correction or spell-check to text-entry fields.

Acknowledgments

The project described was supported by Grant Number 1C1CMS331001-01-00 from the Department of Health and Human Services, Centers for Medicare & Medicaid Services. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services or any of its agencies. The research presented here was conducted by the awardee. These findings may or may not be consistent with or confirmed by the independent evaluation contractor.

Address for correspondence

Lincoln Sheets, MD, MS-CIS
MU Informatics Institute, EBW 248
Columbia MO 65211
sheetslr@health.missouri.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-970

Digital Inclusion for Older Adults based on Physical Activities: an Age Concern

Cristine Gusmão[a,b], Júlio Menezes Jr.[a], Carmelo Pina[c], Juliana Lima[a,b], João Barbosa Neto[a,b]

[a] *SABER Tecnologias Educacionais e Sociais, Universidade Federal de Pernambuco*
[b] *Departamento de Engenharia Biomédica, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco*
[c] *Prefeitura da Cidade do Recife*

Abstract

Nowadays, we are living in an interdependent and interconnected world during an age that is driven by technological progress. It has extraordinary potential to improve the quality of later life: creating social networks to tackle isolation and loneliness; transforming services to help people live independently at home for longer; empowering consumers; and enabling civil participation. In light of this, this poster aims to present the development process of a digital booklet for mobile devices – smartphones and tablets that illustrate the benefits of doing physical exercises for older adults aiming to improve life quality and minimizing digital exclusion.

Keywords:

Older Adults; Physical Exercises; Digital Inclusion.

Introduction

The main barrier of using computers and the internet among older people appears to be a lack of understanding and confidence with 'how it works'. People struggle to comprehend how to use the actual equipment and require explanation as to 'what to press and when' [1].

The benefits of physical activity practice are not restricted to physical-functional and mental field of individuals, but also to social, improving functional performance, maintaining and promoting the independence and autonomy of the older people. Especially among the elderly, it has been found that physical activity decreases the risk of institutionalization and the use of health services and medications [2].

Despite a low level of understanding about how information and communications technology (ICT) can be used, knowledge of its potential benefits is surprisingly high in older people population. From this perspective, this poster presents a project about the development of a digital booklet that will be used to allow older people technological initiation and digital inclusion approaching information about the gains of physical exercise.

Materials and Methods

All methodological actions were planned for integration with extension activities to academic proposals and involved a group of students (undergraduate and posgraduate).

1. Literature review: for understanding the state and progress of current literature on older people health and well being by organizing, integrating, and evaluating previously published books and articles, specially from Brazilian Ministry of Health.

2. Practical habilities: visits were made to the older people group who practice physical activities under Professor Carmelo Pina's supervision. The group consisted of about 50 elderly women, 60 to 90 years, who are under regular medical supervision. The main goal of this phase was to integrate the theoretiical versus practical concepts. All the students attended the activities and contributed to a technical report.

3. Questionnaire application: the questionnaire was applied to a group of 30 elderly women. The main results were: (i) average age: 68,7; (ii) previously read booklets (26); (iii) information gathering location: newspapers (18), magazines (10), and internet (13); (iv) how difficult was it to use new technologies: very easy (1), easy (8), medium (9), hard (10), or very hard (3); (v) internet usage: daily (6), weekly (10), biweekly (2), monthly (1), or almost never (5); (vi) favorite colors: blue and red; (vii) not favorite color: black.

4. Booklet development and testing: the process involved: (i) booklet content writing: after analysis of the group, we selected the following subjects: Mobility, Resistance, Balance and Motor Coordination. (ii) graphic design: using the questionnaires results for the definition of booklet layout; and (iii) user interfaces and application engine: development of the digital booklet through web technologies (HTML5).

Partial Results and Ongoing Work

The use of mobile technologies universalizes and stimulates people's access to health practices. In this case, the elderly population for which it is to exercise more, greatly improving their quality of life. Thus, through the application of educational, older people will be presented with the opportunity to: (i) review concepts and fundamentals of situations and teaching and learning conditions, inserting them and acclimatizing them in the current and future reality of the elderly; and (ii) work the social, psychological, and health (physical and mental) of the elderly.

References

[1] World Health Organization. Global Recommendations on Physical Activity for Health, WHO Press Geneva, Switzerland 2010.

[2] Santos LD, Teixeira-Salmela LF, Lelis FO, Lobo MB Eficácia da atividade física na manutenção do desempenho funcional do idoso: uma revisão da literatura. Fisioter Bras 2001; 2:169-76.

Address for correspondence

Cristine Gusmão (cristinegusmao@gmail.com) Espaço SABER, sala 04, Centro de Convenções, Universidade Federal de Pernambuco – UFPE. Av. dos Reitores - Cidade Universitária, Recife - PE, Brazil. Zip code: 50741-000.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-971

Development of the eHealth Literacy Assessment Toolkit, eHLA

Dorthe Furstrand[a, b], Lars Kayser[a]

[a] *Department of Public Health, University of Copenhagen, Denmark*
[b] *Danish Cancer Society Research Center, Denmark*

Abstract

In a world with rising focus on the use of eHealth, the match between the competences of the individual and the demands of eHealth systems becomes increasingly important, thus making assessment of eHealth literacy as a measure of user competences a vital element.

We propose the eHealth Literacy Assessment toolkit, eHLA, evaluating the user by seven scales: computer familiarity, confidence, incentive and performance as well as functional health literacy, health literacy self-assessment and health literacy performance, as a first step toward development of technology that accommodates the literacy level of the user.

Keywords:

Self-Assessment; Ehealth; Telemedicine; Computer Literacy; Health Literacy; Information Literacy; Patient Participation.

Introduction

Digital health services have become an integrated part of modern healthcare, exposing a growing range of individuals to the challenges of using and benefitting from such systems. For eHealth to be valuable, it is imperative to look at the match between the abilities of the intended user and the skills needed for relevant use of the technology. If a mismatch is to be met, an assessment of user competences, as embraced by the concept of eHealth literacy, must be made. The objective of this project was to develop an assessment tool for this purpose. Health literacy is described as the ability to manage health information in order to maintain and promote good health [1], and eHealth literacy expands the term to include the digital aspect of seeking, finding, understanding, appriasing and applying health information [1]. The Lily Model of eHealth literacy as proposed by Norman and Skinner in 2006 [1] defines the concept by six sub-literacies: traditional, information, media, health, computer and scientific literacy, together providing a frame for describing many of the aspects of competences needed. Since this model, the way we use digital services has changed, and communication, collaboration and content generation takes a larger part. The model of eHealth Literacy by Norgaard [2] is created through a grounded approach taking in both the user and the system perspective. Five out of seven domains hold the view of the user; including knowlegde about one's own health, ability to interact with information and with technology, drive to engage with technology and feeling in control and secure when using technology.

Methods

The foundation of eHLA is an appraisal of predominant methods of assessing health literacy and digital literacy, and the aspects evaluated. eHLA originates from a toolkit used by Knudsen et Vang [3] for eHealth literacy and usability assessment of COPD patients, and was expanded to map up to the Lily model and the Norgaard model, as frameworks of eHealth literacy. The iterative development process included workshops, cognitive interviews and a small scale pilot test of the toolkit.

Results

The seven tools included in eHLA, as seen in Table 1, take care to evaluate self-assessed competences, drive for engagement and comfort, as well as performance, thus sculpting the toolkit around the two models.

Table 1 – Tools included in eHLA

The 7 eHLA tools
Digital literacy tools
Computer Familiarity
Computer Confidence
Computer Incentive
Computer Performance
Health literacy tools
Functional Health Literacy
Health literacy Self-assessment
Health Literacy Performance

Conclusion

The eHLA toolkit showed promising preliminary results for all but one tool, Computer Performance, making it ready for further testing and adjustment, which is now underway.

eHLA and eHealth literacy assessment will enhance the knowledge of user abilities, thereby aiding digital health services in being applicable and beneficial to all patients, thus contributing to the bridging of the social divide in health.

References

[1] Norman CD, and Skinner H. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *J Med Internet Res* 2006: 8(2): e9.

[2] Kayser L, Osborne RH, Norgaard O, Kushniruk A, and Turner P. E-sundhedskompetence - et redskab til at skabe bro mellem borgere, patienter og vores digitale sundhedstilbud. In: eSundhedsobservatoriet 2014, 2014.

[3] Knudsen AK, and Vang AR. Fokus på e-health literacy ved hjemmemonitorering af KOL-borgere. University of Copenhagen, 2012.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-972

Type 1 Diabetes in Twitter: Who All Listen To?

Elia Gabarron[a,b], Alexandra Makhlysheva[a], Luis Marco[a,b]

[a] *Norwegian Centre for Integrated Care and Telemedicine, UNN, Tromsø, Norway*
[b] *Faculty of Health Sciences, Department of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway*

Abstract

Knowing what the conversation on Twitter regarding type 1 diabetes (T1D) is about can help in understanding the kind of information relevant to the individuals affected by the disease. The profile of Twitter users posting on T1D was collected and classified. The number of re-tweets was also registered. The tweets posted by non-governmental organizations (NGOs), communication media, and individuals affected by T1D had higher number of potential readers. More than a half of the tweets were posted by individuals affected by T1D, and their tweets were the most re-tweeted. The next most active users were NGOs and healthcare professionals. However, while tweets soliciting for research funds posted by the NGOs were the next most re-tweeted messages, tweets posted by healthcare professionals were the least re-tweeted. Twitter could be used more actively by healthcare professionals to disseminate correct information about T1D.

Keywords:
Type 1 Diabetes Mellitus; Social Media; Twitter Messaging.

Introduction

Prevalence studies estimate around 500,000 children aged under 15 years have type 1 diabetes (T1D) worldwide [1]. These children and their parents might be looking for information on the disease online, especially on social media, since it has become one of the first sources of information about health [2]. As other diseases, T1D is one of the widely discussed topics on social media.

The aim of this study is to describe the profile of Twitter users posting about T1D, and the potential relevance of the tweets, based on the number of re-tweets and followers.

Materials and Methods

In December 2014, we searched Twitter to sample in total 300 random tweets containing the hashtags #type1diabetes, #t1d, or #type1. Information regarding the tweets emitters, as well as the number of re-tweets were collected. The tweet emitters were classified into six categories: individuals affected by T1D (either patients or parents of children with T1D); healthcare professionals; non-governmental organizations (NGOs); information websites; journalists or communication media; and private companies.

Results

The 300 downloaded tweets on T1D were potentially read by 549,676 readers, according to the number of followers of the emitters at the moment the tweets were posted. Tweets posted by NGOs, communication media and individuals affected by T1D had higher number of potential readers: 210,086; 122,232; and 90,084 respectively.

Most of the tweets from individuals affected by T1D were personal experiences or reccomendations; while tweets from NGO solicited funds for research. Details on the number of tweets per type of emitter, as well as the frequency of re-tweeted messages are summarized in Table 1.

Table 1 – Tweets and re-tweets by user profile

Tweet emitter nature	Tweets on T1D	Re-tweets
Individuals affected by T1D	160 (53.3%)	241 (47.8%)
Non-governmental organizations	37 (12.3%)	139 (27.6%)
Healthcare professionals	36 (12.0%)	14 (2.8%)
Private companies	32 (10.7%)	46 (9.1%)
Communication media	19 (6.3%)	36 (7.1%)
Information site	16 (5.3%)	28 (5.6%)
TOTAL	**300 (100%)**	**504 (100%)**

Conclusion

Individuals affected by T1D are the main group of Twitter users posting about the disease. The personal experiences and reccommendations shared by individuals affected by T1D seems to be the most interesting topic for other Twitter users. The next most active users were NGOs and healthcare professionals. However, while tweets posted by NGOs soliciting research funds were the next most re-tweeted messages, tweets posted by healthcare professionals were the least re-tweeted. Twitter could be used more actively by healthcare professionals to disseminate correct information about T1D.

References

[1] International Diabetes Federation. IDF Diabetes Atlas, 6th ed. 2013.

[2] Jain AV, Bickham D. Adolescent health literacy and the Internet: challenges and opportunities. Curr Opin Pediatr 2014; 26(4):435-439.

Address for correspondence

Elia Gabarron; Norwegian Centre for Integrated Care and Telemedicine (UNN) Sykehusvn. 23, 9019 Tromsø

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-973

Characterizing Consumer Health Informatics in Low and Middle Income Countries

Tugba Kutun[ab], Miriam Föller-Nord[a], Thomas Wetter [bc]

[a] Dept of Computer Science, Mannheim Univ of Appl Sci, Germany, [b] Inst for Medical Biometry and Inform., Heidelberg Univ, Heidelberg, Germany, [c] Dept Biomed Inform and Medical Educ, Univ of Washington, Seattle (WA) USA

Abstract

Consumer Health Informatics (ConsHI) involves patients in health care through ICT, with Low and Middle Income Countries recently entering the field. Compelling successes and complete failures call for the identification of success factors. Of 1092 automatically retrieved articles, 85 were classified as ConsHI. Their service characteristics and the economic and societal factors of the countries of origin were classified. Descriptive statistics were applied in the search for clusters of features that together appear as driving factors. Most factors (financial endowment, number of languages spoken etc) showed no or paradoxical effects. Societal maturity and low population density appear as enabling factors.

Keywords:
Consumer Health Informatics, Low and Middle Income Countries

Introduction

ConsHI offers preventive or medical services through ICT and patient empowerment, driven by the shortage of clinical workforce, opportunities to reach underserved populations at distant locations and the benefits of patient buy in. In a 2014 investigation [1] we looked at the expectations for ConsHI in Low and Middle Income Countries (LMIC). Drawing on paradigmatic examples, we suggested project-internal and environmental success factors. [1] pointed to tentative factors now addressed more broadly.

Material and Methods

PubMed articles were searched with a "Core" process as in [1] for *SMS, cell phone,* etc. and then manually sifted. For active countries with >= 1 confirmed finds, we added a "Ceiling" search with advanced tech terms *Internet, smart phone,* etc. and checked the sifted finds for features (medical problem, service type). Active countries' features were taken from the World Bank web site. A 5-tier classification of the 125 poorest countries (per capita Gross Domestic Product, pcGDP) was applied as in [1] (tier 1 = poorest). In tier analysis, hypotheses for low activity in tiers 3/4 were tested. Next, correlations between numbers of articles and country features were calculated using ranks (r) for clearly non-Gaussian features.

Results

Of 1092 retrieved articles, 85 were considered ConsHI in Core or Ceiling. 31 from India and China are excluded in Table 1b due to their huge populations' distorting effects, leaving 24 countries. A paradox from [1] of lowest activity in tiers 3 and 4 is confirmed for Core (Table 1a). Adding Ceiling makes 3 the single lowest. 7 countries in tiers 3/4 are one of the 15 former USSR republics. Table 1b shows the strongest correlations with article numbers in PubMed (subsequently: Activity). The negative correlations insinuate that low geographical latitude and low population density are enabling factors.

Table 1a – Activity by tier

Tier	Core	Ceiling
5	16	2
4	9	13
3	5	0
2	25	1
1	11	3

Table 1b – Activity vs feature

Country Feature	Corr. Coeff.
Distance to equator	-0.36
Internet users (r)	0.36
Female literacy	0.30
Population density	-0.27
Languages	0.26

Table 2a – Service types

Service type	Core	Ceiling
Prevention	21	2
Therapy	17	12
Follow up	11	0
Education	4	8

Table 2b – Medical problems

Condition	# Art.
HIV	28
Mental disorder	11
Bacteria/Parasites	9
Diabetes mellitus 2	9

Prevention ranges from dental hygene to safe sex. Many therapies are drug reminders. Depression and suicide lead the Mental list. Richer media in Ceiling benefit Education (Table 2).

Discussion

The small sample size limits the validity of conclusions and the results resist easy explanations. Closeness to the equator and thus attracting development aid suggests itself. More widely, if money helps, International investment or Increase in pcGDP should correlate with Activity. Number of cell phones should also correlate, but neither does. Internet usage, Female literacy and Tourism suggest a societal maturity "Gestalt". Low population density fosters services that cover distances. Method as in [2] may substantiate such claims.

Conclusion

The lack of common success factors found here supports the service-to-context fit proclaimed in [1] as prevailing factor.

References

[1] Hartzler A and Wetter T Engaging Patients through Mobile Phones: Demonstrator Services, Success Factors, and Future Opportunities in Low and Middle Income Countries. Yearb Med Inform 2014: 182-94

[2] Séror AC A Case Analysis of INFOMED: The Cuban National Health Care Telecommunications Network. J Med Internet Res 2006: 8(1): e1

Address for correspondence

Thomas.Wetter@urz.uni-hd.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-974

3D CPR Game Can Improve CPR Skill Retention

Jia Li[a], Yimin Xu[a], Yahong Xu[a], Peng Yue[a], Liu Sun[a], Ming Guo[a], Shuqin Xiao[a], Shu Ding[a], Yanyan Cui[a], Shulan Li[a], Qiuying Yang[b], Polun Chang[c], Ying Wu[a,*]

[a] *School of Nursing, Capital Medical University, Beijing, China*
[b] *School of Biomedical Engineering, Capital Medical University, Beijing, China*
[c] *Institute of biomedical informatics, National Yang-Ming University, Taiwan*

Abstract

Adequate cardiopulmonary resuscitation (CPR) skill is essential in improving survival rate of sudden cardiac arrest (SCA). However, the skill deteriorates rapidly following CPR training. We developed a computer game by using 3-Dimensional virtual technology (3-D CPR game) for laypersons in the purpose to improve skill retention. As the testing phase, a randomized control trial, in which we recruited 97 freshman medical students who had no prior CPR training experience, was used to test its effect on 3-month CPR Skill retention. The usability of the game was also tested using a 33 item questionnaire rated with 5-point Likert scale. Three months after the initial CPR training, the retention rate of CPR skill in the game group was significantly higher compared with the control ($p<0.05$) and the average score on 4 dimensions of usability were 3.99-4.05. Overall, using 3-D CPR game in improving CPR skill retention is feasible and effective.

Keywords:

CPR; 3-D computer game; skill retention; Cardiac arrest

Introduction

SCAs are usually occurred in out-of-hospital settings (OHCA). Bystander CPR is an effective way of improving the outcome of OHCA. However, the survival rate depends on the quality of bystander's CPR skill which could deteriorate quickly even at 2 weeks after training [1]. Although our previous study showed that remote supervised reinforcement of CPR skills by telephone could improve CPR skill retention, adherence was poor. Because game can provide a learning-by-playing environment and motivate trainees, we designed a 3-D CPR game for laypersons to refresh CPR skills and improve skill retention. This pilot study was to test the effect and the usability of the game by recruiting freshman medical students who had no prior CPR training experience, therefore provide evidence for further usage of CPR game in improving skill retention among a larger population of laypersons.

Materials and Methods

3-Dimensional virtual technology and real-world scenario were used to develop a CPR computer game. After repeated modification and trial use, freshman medical students who had no prior CPR training experience were recruited and randomly assigned into game or control group using sealed envelopes if they pass the test after AHA basic life support *Heartsaver* course. Subjects in the game group were instructed to play the game to approach a victim and provide resuscitation according to the CPR skill they learned, while control group were told to refresh their skills frequently. Three months (±2 weeks) following the CPR training, all the subjects who had passed the initial examination were invited to return for re-examination on Resusci Anne manikins which can printout data on volume of ventilation, rate and depth of compression, and compression site. The usability of the game was also evaluated on a 5-point Likert scale questionnaire, which was developed according to Technology Acceptance Model of Davis, F. D. (1989). The content validity for the total scale was 0.95 and the Cronbach's alpha was 0.98.

Results

The interfaces of the final CPR game are shown in Figure.1. There are two avatars (one SCA victim and one rescuer). The player can control the rescuer avatar by computer mouse and keyboard to act as a rescuer to perform CPR as learned during the CPR training course.

During the testing phase, 97 subjects were recruited and finished the study. The 3-month retention rates of overall performance, compression depth, airway opening, and ventilation volume were higher in the game group compared to the control ($p<0.05$). The average score on 4 dimensions of usability scale for this game were 3.99-4.05.

Figure.1 - Resuscitation game interfaces

Conclusion

Using 3-D CPR game in improving CPR skill retention is feasible and effective among freshman medical students who had no prior CPR training experience, and the effect for laypersons needs to be further investigated in a wider range of population.

References

[1] Roppolo LP, Saunders T, Pepe PE, Ldris AH. Layperson training for cardiopulmonary resuscitation: when less is better. Current Opinion In Critical Care. 2007;13(3):256-60.

Address for correspondence

∗ Corresponding author. Ying Wu.
E-mail address: helenywu@vip.163.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-975

Improvement of Hemoglobin with Repeated Health Checks Among Women in Bangladesh

Mariko Nishikitani[a], Yasunobu Nohara[b], Partha Ghosh[c], Rafiqul Islam Maruf[c], Ashir Ahmed[a,d], Naoki Nakashima[a,b]

[a] *Institute of Decision Science for a Sustainable Society, Kyushu University, Fukuoka Japan*
[b] *Medical Information Center, Kyushu University Hospital, Fukuoka, Japan*
[c] *Grameen Communications, Dhaka, Bangladesh*
[d] *Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan*

Abstract

The residents of several cities and villages in Bangladesh underwent e-health checkups. The distribution of Hb increased from the first hemoglobin (Hb) measurement to the second, and some women recovered from severe anemia. A strong effect was observed in subjects who were prescribed iron supplements. The mean blood pressure improved after the first health examination, and the blood sugar level increased. In countries with shortages of health practitioners and health care providers, e-health examinations and instructions may be useful interventions because they make people health conscious and improve their health status.

Keywords:

Preventive medicine; E-health; Health examination; Hemoglobin (Hb); Women; Developing countries

Introduction

In developing countries, low hemoglobin (Hb) suggests inferior health and nutritional status and indicates risks for increased maternal and child mortality. If Hb information is provided it may help prevent disease development and death.

In this study we checked Hb level as part of a general health examination in Bangladesh. The objective of this study was to assess the effect of the health examination and the subsequent health instructions on Hb levels in women.

Methods

E-health checkups were provided for the residents of Dhaka, five villages, and employees in several factories/offices in Bangladesh during 2012 and 2013. Individual health measurements were recorded automatically and diagnosed based on international diagnostic standards. Tele-consultation and tele-prescription were provided [1].

The first checkup at several venues of six cities in 2013 included 10,575 subjects (5,043 women, 48%). The study subjects were the 1,205 women who had abnormal Hb levels at their first checkups. Two months later, a second checkup was conducted in five cities and some venues, and 348 women returned.

Results

The mean baseline Hb of the returning group (n = 348) was 10.8 ± 1.7 g/dL and it did not differ significantly from that in women seen only at the first visit (n = 857, 11.0 g/dL). The mean Hb at the second checkup among the returning group increased to 11.5 ± 1.3 g/dL (Figure 1). Blood pressure decreased by 4–5 mmHg. Body mass index (BMI) did not differ significantly between checkups, and blood sugar increased significantly when comapred with the first visit.

To confirm the effect of health instructions following the first checkup, several interventions were assessed, including suggested prescription drugs, prescribed iron (Fe) supplements, dietary instructions, and referrals to specialists. The standardized partial regression coefficient adjusted for age, job, and place of checkup showed that the prescription of Fe supplements had a significant positive effect on Hb (beta = 0.325, $p < 0.001$).

Conclusion

Repeated health examinations and subsequent health instructions were useful interventions that improved Hb and other health indicators among women in Bangladesh.

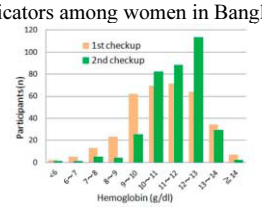


Figure 1 – Change in the distribution of measured Hb

References

[1] Nohara Y, Kai E, Ghosh PP, Islam R, Ahmed A, et al. Health Checkup and Telemedical Intervention Program for Preventive Medicine in Developing Countries: Verification Study. J Med Internet Res 2015: 17(1);e2.

Address for correspondence:

CollaboII-122, 3-1-1 Maidashi, Higashi-ku Fukuoka 812-8582, Japan. makorin@med.kyushu-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-976

Cardiac auscultation simulator embedded in virtual learning environment to support medical teaching

Isaias Oliveira[a], Zilma Silveira N Reis[a], Marilene M Araújo[b], Claudia Maria V Freire[b]

[a] School of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[a] Prenatal care service for cardiac diseases, Hospital das Clínicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract

Heart auscultation remains as an important clinical tool in heart disease diagnostics. A prototype was developed, using an existing database of clinical and cardiac sounds of pregnant women [1]. The context for virtual learning is cardiac disease during pregnancy (Partum and Post Partum situations). Based on real histories, cardiac sounds, and complementary exams, this pilot presents the most frequent diagnosis of cardiac diseases during pregnancy.

Keywords:

Heart Auscultation; Digital Stethoscope; Cardiovascular Pregnancy Complications; Medical Informatics, Teaching Materials.

Introduction

Heart auscultation is a highly important clinical tool in diagnosing hearts diseases [2]. However, this detection is still a challenge for professionals who work directly in prenatal health care. It is estimated that most internal medicine and family practice physicians are not able to recognize common heart sounds and murmurs that could be diagnosed using a stethoscope [3].

This study presents a cardiac auscultation simulator using virtual learning environment.

Methods

The simulator was projected to support teaching of cardiac auscultation. The institution's Ethics Committee approved the study, and all volunteers signed a letter of consent.

From the methodological approach of problem-based learning, real clinical patients are the starting point. The purpose is intended to cause the student to be challenged, and therefore active in the learning process, encouraging them to study the desired content. The interface was developed to be intuitive and self-explanatory, with minimal adaptation time.

Results

The system model has a main screen with a female thorax (Figure 1), data collection, and an interactive area interface. Heart sounds sequence took place in the classic areas of auscultation.

While listening to the audio, the student can choose to view patient data. The challenge for students is to reach the true diagnosis of a cardiac disease, choosing from a list of four options.

Discussion/Conclusion

Simulation environments can contribute by allowing repetitive and individualized oriented training in support of learning in real scenarios [4].

In this work, we present a functional prototype system of learning to provide training for normal and abnormal cardiac auscultation.

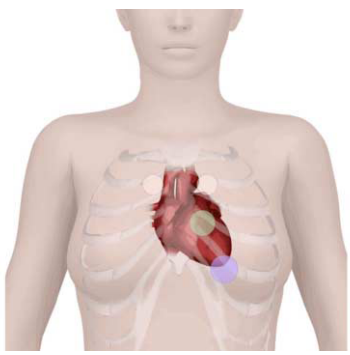


Figure 1 - Female chest and classical points of auscultation

References

[1] Araujo MM et al., inventor; Universidade Federal de Minas Gerais. Sonoteca cardíaca. Instituto Nacional de Propriedade Intelectual. BR 51 2014 000282 9. 2014 Março 21.
[2] Chizner MA. Cardiac auscultation: rediscovering the lost art. Current problems in cardiology. 2008;33(7):326-408.
[3] Mangione S, Nieman LZ. Cardiac auscultatory skills of internval medicine and family practice trainees. A comparison of diagnostic proficiency. JAMA : the journal of the American Medical Association. 1997;278(9):717-22.
[4] Troncon LEA. Utilização de pacientes simulados no ensino e na avaliação de habilidades clínicas. Medicina (Ribeirao Preto Online). 2007;40(2)

Acknowledgments

Supported by FAPEMIG

Address for correspondence

isaias@medicina.ufmg.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-977

Global Challenges in People-Centered E-Health

Yuri Quintana, Charles Safran

Division of Clinical Informatics, Beth Israel Deaconess Medical Center and Harvard Medical School

Abstract

People-centered health care seeks an active role for the patient while empowering all other members of the health care team. By promoting greater patient responsibility and optimal usage, patient-centered health care leads to improved health outcomes, quality of life and optimal value for health care investment. This paper reviews some definitions of people-centered health care and various e-health approaches around the world used to implement this vision. The barriers and enablers to implementation this type of approach are explored. This paper provides a proposed research agenda for future implementations of people-centered e-health.

Keywords:

Person-centred health care; participatory health; health services research; individualized care; complex systems; transdisciplinary research; people-centered e-health.

Introduction

It has been long argued that "the largest and yet least used health care resource, worldwide, is the patient or prospective patient" [1-4]. In recent years, technology has helped to provide patients a greater voice in how health care is accessed and delivered. People-Centered Health Care medicine [5] is a model of medical care in which the active role of the patient is emphasized. The overall vision for people-centered health care is one in which individuals, families, and communities can participate in trusted health systems that respond to their needs in humane and holistic ways. People-centered health care is an umbrella term that better encapsulates the foremost consideration of the patient across all levels of health systems. The definition of people-centered health care has shifted over the years, and the implementation has also shifted; however, the core values have remained the same.

Methods

We searched the PubMed using the search query ("people-centered" or "patient-centered" or "people-centred" or "patient-centred" or "participatory medicine") and ("patient empowerment" or "patient participation" or "family engagement" or "social justice") in titles and abstract, searched the Internet with the same terms. We then reviewed websites and publications and summarized the objectives, approaches and outcomes of some of the frequently cited programs in the areas of patient empowerment, patient participation, family engagement, and social access.

Conclusions

People-centered health care programs [1-9] provide an opportunity to promote patient empowerment, patient participation, family engagement, and social justice. Some programs have shown some promising results by showing increasing numbers of patients using the programs but the evidence on outcomes is still limited. Most of the programs are in developed countries. Some international guidelines have been developed for these programs. This paper has proposed some strategies for the design and implementation of programs in developing countries. More research is needed for how to design and evaluate these programs in different settings and cultures. Patient representatives should be part of product design teams, health care provider committees, and in government policy committees so that programs that are created support patient priorities.

References

[1] Slack WV. Patient counselling by computer. Zoog S. Yarnall S, eds. The changing health care team. Medical Communications and Services Association, 1976, 108-11.

[2] Slack WV. The patient's right to decide. Lancet 1977; 2: 2

[3] Slack WV, Safran CS, Kowaloff HB, Pearce J, Delbanco TL. A computer - administered health screening interview for hospital personnel. MD Comput 1995; 12: 25–30

[4] Safran, C. AMIA President Testimony, Healthcare Information Technology Subcommittee, US House of Representatives, June 17, 2004 http://waysandmeans.house.gov Accessed Dec 16, 2014.

[5] World Health Organization (Western Pacific Region). People-Centred health care - A policy Framework 2007.

[6] Frydman, Gilles. A Patient-Centric Definition of Participatory Medicine. Mar 17, 2013, http://www.patientdriven.org/2013/03/a-patient-centric-definition-of-participatory-medicine/. Accessed December 16, 2014

[7] Ferguson, T. et.al. e-patients: how they can help us heal healthcare. Originally published in the Journal of Participatory Medicine in May 2007. Afterward in 2013. http://e-patients.net/e-Patients_White_Paper.pdf. Accessed December 16, 2014

[8] Society of Participatory Medicine (SOP) 2014 Homepage. http://participatorymedicine.org/.Accessed December 16, 2014.

[9] Lynn Nielsen-Bohlman L, Panzer AM, and Kindig DA (Eds). Health Literacy: A Prescription to End Confusion. Washington, D.C.: National Academies Press, 2004.

Address for correspondence
Yuri Quintana, Ph.D., Director for Global Health Informatics, Division of Clinical Informatics, Beth Isreal Deaconess Medical Cenrter, and Harvard Medical School

Email: yquintan@bidmc.harvard.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-978

Use of Patient Portals: Personal Health Information Management in Older Adults

Anne Turner[ab], Katie Osterhage[b], Jonathan Joe[a], Andrea Hartzler[c], Lorelei Lin[b], George Demiris[a]

[a] *Biomedical Informatics & Medical Education*
[b] *Health Services, University of Washington, Seattle, WA USA*
[c] *Group Health Research Institute, Seattle, WA USA*

Abstract

The personal health information management (PHIM) of older adults is poorly understood. We describe initial results from the SOARING (Studying Older Adults & Researching Information Needs and Goals) study at the University of Washington, a particpatory design investigation of PHIM in older adults. We conducted in-depth interviews with older adults (n=71) living in a variety of residential settings. A surprising 21% of participants reported using patient portals. Another 17% of participants reported prior use or anticipated use of portals in the future. We identified preferences and needs as well as barriers and facilitators to portal use. Our findings indicate that patient portals designed to target the specific needs for older adults can suport PHIM. We offer recommendations for expanded research.

Keywords:

Patient portals; Older adults; Health information management.

Introduction

Older adults have been slow to adopt technologies such as patient portals for managing their personal health information (PHI). A patient portal is a secure website through which patients access information from an electronic health record. These technologies are largely designed without consideration of older adult needs and preferences. Understanding the barriers and uses of technology by older adults for PHIM can inform design imporvements. In the context of a broader project to investigate PHIM in older adults, we report on general computer use and the use of patient portals among older adults across different residential settings.

Methods

We conducted 71 in-depth interviews with older adults aged 60 years (yrs) and recruited from adult residential centers, assisted living, and independent residences in King County, WA USA. We used purposive sampling to ensure diverse representation of gender, socio-economic status (SES), and racial backgrounds. The interviews assessed participant demographics, health, social networks and technology use, followed by open-ended questions regarding PHIM practices. We transcribed and coded interviews for emergent themes.

Results

Participants described PHIM practices ranging from complex systems of medication tracking to discarding all but the most critical health information. Of all participants, 65% (46/71) reported using a computer at least 2-3 days/week, and a majority, 79% (56/71), reported access to the Internet. A portion of participants, 21% (15/71), mentioned current use of patient portals, while another 17% (12/71) of participants had previously used a portal, used a portal through a family member, or indicated potential future use. Current portal users ranged from age 61 to 93 yrs, most lived independently, and responses indicated middle to high SES. The majority felt positive about the specific portal they used, which were primarily versions of Epic MyChart™ (Epic Systems Corp, Verona, WI). Several participants mentioned a dramatic reduction in their health information record keeping because of the portal. One participant reported disconituation of the portal due to log-in frustrations.

Discussion

Despite reports of barriers to the use of PHI technologies by older adults, 21% of the older adult participants use patient portals to manage personal health information. This trend is encouraging for the future. However, the small number of participants who use portals (n=15) in this study limits our ability to analyze the effect of age and living situation on portal adoption. Expanded research is needed to determine the general penetration of patient portals in this population, factors that contribute to portal use by older adults, and associations between portal use, age, and living situation.

Conclusion

Study findings highlight the small, but substantial, role of patient portals as a platform to facilitate management of PHI. The use of patient portals demonstrates their potential role in helping older adults to maintain wellness and independence.

Acknowledgments

This research was funded by the Agency for Healthcare Research and Quality (AHRQ) grant # R01 HS022106. The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ.

Address for Correspondence

Anne Turner, amturner@uw.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-979

Machine Assisted Translation of Health Materials to Chinese: An Initial Evaluation

Anne M. Turner[a], Loma Desai[a], Kristin Dew[a], Nathalie Martin[a], Katrin Kirchhoff[b]

[a] *Department of Biomedical Informatics and Medical Education*
[b] *Department of Electrical Engineering, University of Washington, Seattle, WA, USA*

Abstract

There is an unmet need for Chinese language health materials in the USA. We investigated the use of machine translation (MT) plus human post-editing (PE) to produce Chinese translations of public health materials. We collected 60 documents that had been manually translated from English to traditional Chinese. The English versions were translated to Chinese using MT and assessed for errors and time required to correct via PE. Results suggest poor initial translation may explain the lack of quality translations despite PE.

Keywords:
Public health informatics; natural language processing; machine translation, Chinese language, health promotion.

Introduction

Chinese is the second most common language among limited English proficient (LEP) individuals in the USA. LEP status has been associated with negative health consequences, including poorer health outcomes and poorer access to health care. Unfortunately, due to the time and costs of producing quality translations, very few health materials are available in languages other than English. Our previous research indicates that free online translation engines, such as Google Translate, can be used in conjunction with human post-editing (PE), to produce quality English to Spanish translation quickly and at low cost. To investigate the usefulness of MT in translating health materials for a more difficult language pair, we investigated the types of errors in English-to-Chinese machine translations, the PE time to correct them, and the quality of MT plus PE compared to manual translation.

Methods

We collected 60 health promotion documents from various public health websites and conducted three stages of experiments with them: linguistic error analysis, PE, and quality rating. Five native Chinese participants with health experience performed PE. We recorded the time taken to PE each document. Two bilingual public health professionals then independently, blindly rated a subset of 20 pairs of manually translated and MT-plus-PE documents to indicate which they preferred and why.

Results

The linguistic error analysis found that word sense errors were most common (40%), followed by word order errors (22%) and missing words (16%).

PE time as measured by the mean charaters/min CPM per document varied from 18.5 to 79.6 CPM (SD = 0.03-38.7). The mean CPM across all documents was 38.9 (SD = 9.9); thus, on average a post-editor can correct approximately 39 CPMs with a variation of around 10 CPMs. There was no relationship between document length and post-editing time. PE participants rated the adequacy of the MT output at 3.6, suggesting that "much" to "most" of the original meaning of the source text was preserved with MT. The average fluency was rated 3.4, which indicates a level of grammar quality between "non-native Chinese" and "good Chinese". There were large differences in PE speed between post-editors, ranging from 18 to 41 CPM.

In the blind comparison of MT plus PE and manual human translations, both quality raters preferred the manually translated document in all 20 cases. Reasons given for the preference were better word order, professional reading level, flow, word accuracy, and cultural appropriateness.

Discussion

Although our prior research using English to Spanish translation indicated that MT with PE could produce a translation equivalent in quality for less time and cost, translation from English to traditional Chinese showed that maintaining quality through post-editing was more problematic. The final post-edited translations still contained errors that made the quality rater prefer manual translations. There are several possible reasons for this difference including: differences in MT quality for this language pair; differences in instructions provided to post-editors; linguistic expertise of post-editors; engagement of post-editors; and subtleties of cultural appropriateness.

Conclusion

Despite great need, there are few health promotion materials available in Chinese. Our investigation of the use of MT plus PE to produce translations indicated that using the methods that worked for English-Spanish translations were not as effective for English-Chinese. Multiple factors, including the quality of the MT and the expertise of the post-editors, may contribute to this result. We are performing additional studies to determine how best to improve translation from English to Chinese in order to ensure quality translation at low cost.

Address for correspondence

Anne M. Turner, amturner@uw.edu, UW Box 354809, Seattle, WA, USA.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-980

Lessons Learnt from Evaluation of Computer Interpretable Clinical Guidelines Tools and Methods: Literature Review

Soudabeh Khodambashi[a], Øystein Nytrø[a]

[a] *Department of computer and Information Science, Norwegian University of Science and Technology, Norway*

Abstract

Representation of clinical guidelines in a computer interpretable format is an active area of research. Various methods and tools have been proposed which have been evaluated based on different evaluation criteria. The evaluation results in the literature and their lessons learnt can be a valuable learning resource in order to redesign and improve the tools. Therefore, this research investigates the lessons learnt from the evaluation studies. Broad search in literature together with a purposeful snowball method were performed to identify the related papers that report any type of evaluation or comparison. We reviewed and analysed the lessons learnt from the evaluation results and classified them into 17 themes which reflect the suggestion concerns. The results indicate that the lessons learnt are more focused on tool functionalities, integration, sharing and maintenance domain. We provide suggestions for the area which had less attention.

Keywords:

Computer interpretable clinical guidelines; Evaluation; Lessons learnt.

Introduction

Clinical guidelines mostly are written in a free-text format from which it is hard to search and retrieve recommendations during medication in a short period of time. Several improvements with respect to computer interpretable clinical guidelines (CIGs) have been reported [1]. Some of the tools and methods were evaluated based on sets of criteria independently in order to assess the extent to which the tool can satisfy certain aspects [2]. As each of the evaluations and their results are valuable and much can be learned by comparing the findings, the main goal of this research was reviewing the results of evaluations studies and their lessons learnt regarding both technical and sociotechnical aspects in the domain of CIGs.

Methods

The research method was literature review in the domain of CIGs. We searched in the Journal of Biomedical Informatics, Journal of the American Medical Informatics Association, Methods of Information in medicine, International Journal of Medical Informatics, Artificial Intelligence in Medicine and Digital Bibliographic Library Browser to find the relevant papers. The search keywords were "computer interpretable guidelines", "clinical guideline formal representation", "electronic clinical guideline", and "computerized clinical guideline". In addition, a purposeful "snowball" sampling method enabled us to gain more relevant papers which were cited in the full-text of the reviewed papers.

Results

A total number of 27 papers identified as relevant for full-text review. We grouped similar results into 17 thematic categories. The similarities are identified based on their suggestions for improvement in specific domain. The identified themes are: user support in modelling and encoding process, standard terminology and data model, consistency checking, knowledge specification step, guideline content and dimensions, visual support and editor features, variation, ontology and semantic tagging, macros, representation on different level of abstraction, representation primitives, automatic knowledge extraction, concurrent guideline development and formalization, configurable component-based modelling constructs, required skill levels, tool functionality, didactic and maintenance support.

Discussion

The results indicate that lessons learnt are more focused on functionalities that a tool needs to support, as well as integration, sharing and maintenance considerations (i.e. standard terminology and data model). In addition, factors that reduce variability and improve consistency of the encoded guidelines are another area of concern. We see fewer suggestions related to visualization support, especially regarding adaptation to the encoder's skill level. Based on the results, visualization support with provided didactic material that is adapted based on user's skills can improve usability and reduce variability in encoding clinical guidelines.

Conclusion

The results of this paper are valuable for researchers and designers of the tools in the domain of CIGs. To our best knowledge, this is the only broad literature review on the results and evaluation lessons in the domain of CIGs.

References

[1] Ciccarese P, Caffi E, Quaglini S, Stefanelli M. Architectures and tools for innovative Health Information Systems: The Guide Project. International journal of medical informatics. 2005;74(7):553-62.

[2] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A. Comparing computer-interpretable guideline models: a case-study approach. Journal of the American Medical Informatics Association. 2003;10(1):52-68.

Address for correspondence

Soudabeh@idi.ntnu.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-981

The need for cost-benefit analyses of eHealth in low and middle-income countries

Fleur Fritz[a], Mihiretu Kebede[b], Binyam Tilahun[a]

[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Health Informatics, University of Gondar, Ethiopia*

Abstract

Cost and benefit analyses are crucial for eHealth interventions, especially in low and middle income countries (LMIC). We performed a scoping review, with goals to acquire an overview of cost-benefit studies and to collect indicators that are being used for costs and savings with respect to eHealth. Of all retrieved articles, 29 were included in our analysis and 21 contained cost/savings indicators which we categorized into outcome measures with respect to providers, patients and other stakeholders. Nearly all authors stated that more evidence from economic analyses by health economists are urgently needed and that pilot projects should collect cost related data for evaluation.

Keywords: *E-health; cost-benefit; efficiency; low-income*

Introduction

Many LMIC face the challenge of providing medical care with limited resources. During the last decade there has been an increase in the use of health information technology (HIT) to support medical work [1]. However, if LMIC want to empirically decide on which eHealth application is worthwhile for countrywide use, they are largely dependent on small and scattered pilot installations, oftentimes without thorough evaluations. Many of the pilot studies that are being conducted report on the feasibility of the implemented systems, but analyses on impact and cost efficiency are largely missing [2]. Despite this clear deficit, in a setting in which there is insufficient funding to provide adequate medical staffing for basic medical care, the decision to use resources for IT must be well-justified. Our objective in this study therefore is to provide an overview on current research about eHealth interventions in LMIC with respect to cost-benefit analyses.

Materials and Methods

We conducted a scoping review by searching the PubMed database and respective reference lists. In the analysis we collected all direct or indirect mentioned cost and savings indicators. The indicators were mapped according to Schweitzer et al.'s outcome measures for mhealth (originally derived from Davalos et al. [3]) [4].

Results

Our search resulted in 90 articles, of which 29 were included according to our inclusion criteria. Of the 29 articles, 15 were evaluation studies, 10 review papers, 3 lessons learned and 1 use case. Not all analysed articles explicitely named cost and savings indicators. From 21 we were able to extract a total of 50 measures. In a consensus process we removed redundant indicators, while renaming and merging common variables. This process resulted in 32 distinct cost and saving indicators. We then compared the collected indicators with Schweitzer et al.'s measures of clinical and social outcomes from patient, provider and stakeholder perspectives [4].

The results show that eHealth applications primarily benefit healthcare providers in terms of data quality, service improvement and paper cost savings. Second, they benefit other healthcare stakeholders by improving communication, and reporting, and remote patient monitoring. Third, they benefit the patient by reducing travel and and waiting times, in addition to improving diagnostic accuracy and clinical efficiency.

Discussion & Conclusion

Although the number of implementations, and consequently evaluations and review papers on eHealth in LMIC are increasing, evidence on cost-benefit, particularly from a patient's perspective, is lacking. There is an urgent need to perform economic analyses to effectively use the limited available resources in LMIC. A validated framework for economic analyses of eHealth interventions is still lacking. In order to develop suitable methodologies and validate HIT evaluation frameworks in the context of developing countries, the HIT community should join forces with economists.

References

[1] Luna D, Otero C, Marcelo A. Health Informatics in Developing Countries: Systematic Review of Reviews. Contribution of the IMIA Working Group Health Informatics for Development. Yearb Med Inform. 2013;8(1):28-33.

[2] Piette JD, Lun KC, Moura LA Jr, Fraser HS, Mechael PN, Powell J, Khoja SR. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? Bull World Health Organ. 2012 May 1;90(5):365-72.

[3] Dávalos ME, French MT, Burdick AE, Simmons SC. Economic evaluation of telemedicine: review of the literature and research guidelines for benefit-cost analysis. Telemed J E Health. 2009 Dec;15(10):933-48.

[4] Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun. 2012;17 Suppl 1:73-81.

Address for correspondence: Fleur.Fritz@uni-muenster.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-982

Effects of electronic prescription on pharmacy productivity

Reima Suomi[a], Markus Lähteenoja[b] and Sirpa Peura[c]

[a] University of Turku [b] University of Turku [c] Association of Finnish Pharmacies

Abstract

Electronic prescriptions affect pharmacy workflows. In this study, we measure the workflow efficiency in pharmacies in 2006 and 2012: both, in traditional workflow when electronic prescription was not in use, and in new direct delivery workflow, which is the mandated workflow model in the case of electronic prescriptions.

Keywords:

Electronic prescription, pharmacy productivity, retail pharmacy, community pharmacy, outpatient pharmacy, pharmacy workflow

Introduction

Electronic prescriptions have been widely used in Finland during 2012 – 2014. A switch to electronic prescription changes the counter delivery in the pharmacies from a workflow that we call "traditional model" to a new one that we call "direct delivery model". In this research we study what this change means to the productivity of the Finnish pharmacies.

Methods

In the research, we measured the interaction time needed to take care of one prescription at the counter of retail pharmacies. The process started from the customer entering at the counter, and ended with the handout of the medicine. Payment time for the medicine was not included.

Measurements were taken at two timepoints: in years 2006 and 2012.The year 2006 results have been reported elsewhere [1], and are not a topic of this presentation. The detailed results of the 2012 study are reported in [2]. Methodologically our research is quite close to traditional Taylorism.

Results

Detailed data on the delivery processes is in Table 1.

Table 1. *Number of prescription delivery processes analyzed in 2006 and 2012*

	Paper prescription 2006	**Paper prescription 2012**	**Electronic prescription 2012**
Customer sessions	123	347	162
Number of prescriptions handled	190	573	275
Number of packets delivered	234	704	322

Table 2 contains the needed times for medicine delivery.

Table 2. *Total times of delivery for different prescriptions in 2006 and 2012*

	Paper prescription 2006	**Paper prescription 2012**	**Electronic prescription 2012**
median time	2 min 38 s	1 min 43 s	2 min 4 s
average time	2 min 48 s	1 min 58 s	2 min 26 s
standard deviation	1 min 34 s	57 s	1 min 17 s
minimum time	34 s	32 s	36 s
maximum time	12 min 40 s	6 min 22 s	9 min 15 s

In general, medicine delivery process in pharmacies has improved during the timespan of 2006 – 2012, making the prescription handling times faster.

Discussion

E-prescriptions have not increased pharmacy productivity. Time needed to handle an electronic prescription is down by 13% from year 2006 figures. However, the improvement in handling time is even better (30%) for the paper prescriptions.

Conclusion

In the Finnish environment, the results show that with the electronic prescription, the delivery time for a single medication over the counter was cut by 13%. In other words, the pharmacists are able to deliver 10 prescriptions in the same time that previously allowed nine prescriptions to be handled. This Indicates that productivity in pharmacies has grown a lot during the studied time period in Finnish pharmacies.

References

[1] M. Koiranen, and J. Sinnemäki, "Reseptitoimitukseen kuluneet ajat suoratoimitus- ja perinteisellä mallilla (Times needed for prescription delivery with the traditional and direct delivery workflow models) Unpublished manuscript.," Turku School of Economics, 2006.

[2] M. Lähteenoja, "E-reseptin ja paperireseptin toimitusajat suoratoimitusapteekeissa (Delivery times for paper and electronic prescriptions in pharmacies using direct delivery model). Unpublished Master's Thesis.," University of Turku, 2014.

Address for correspondence

Reima Suomi, reima.suomi@utu.fi

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-983

(Br-SCMM) Brazilian Smart City Maturity Model: A Perspective from the Health Domain

Ricardo Alexandre Afonso[a,b], Kellyton dos Santos Brito[a,c], Clóvis Holanda do Nascimento[b], Luciana Campos da Costa[a], Alexandre Álvaro[b,d], Vinicius Cardoso Garcia[b]

[a] *Alagoas Federal University (UFAL), Arapiraca AL, Brazil*
[b] *Informatics Center (CIn) – Pernambuco Federal University (UFPE), Recife PE, Brazil*
[c] *Pernambuco Federal Rural University (UFRPE), Recife PE, Brazil*
[d] *São Carlos Federal University (UFSCar), Araraquara SP, Brazil*

Abstract

The term definition "Smart City" still allows various interpretations, and this causes some difficulty in establishing parameters to measure how smart the cities can be. This paper presents a Maturity Model that uses a set of minimum domains and indicators that aim to encourage cities of different sizes to identify their potential and improve processes and public policies.

Keywords:

Public health; Big Data; Government; Smart Home.

Introduction

The concept of Smart Cities is still confused. Some authors suggest the use of scenarios related to transportation, health, technology, infrastructure, stability and others [1,2]. The purpose of this work is to present a way to measure how smart a city can be. In addition, to propose an incentive mechanism for cities to come to the ranking of the Smart Cities emphasizing the domain Health of these cities; and what indicators can be used to make this comparison between municipalities.

Methods

To create the rating for smart cities, the attention on creating minimum measures which may be affected by these smaller municipalities is needed. Therefore, this work proposes to create a model composed of ten areas called "Domains Basic" where each domain has its respective "Basic Indicator" [3].

Table 1 – Basic Domains and Indicators

Basic domains	***Basic Indicators***
A. Water	Piped water
B. Education	HDI-Education
C. Energy	Access to energy
D. Governance	HDI Employment
E. Housing	Private residence
F. Environment	Garbage collected
G. Health	HDI, ISO37120
H. Security	Homicides per 1000
I. Technology	Computer at home
J. Transportation	Mass transit

Results

The model proposed in this work was called Br-SCMM (Brazilian Smart Cities Maturity Model). It uses the domains and indicators presented in results of this study to measure the first level on a scale from 1 to 5 to identify possible areas for improvement before that the following levels are adopted. The levels are divided into five categories that compose the word (SMART): Simplified, Managed, Applied, Measured and Transformed. All Brazilian capitals were measured using the S level of this model.

Conclusion

This work presented the domains and indicators that will be used to measure the capacity optimization and improvement of municipal resources and processes. The model (Br-SCMM) is being developed in partnership with three Brazilian universities and the municipalities where these institutions are located so that you can implement and test the following levels of this model.

Acknowledgments

This work was supported by the FACEPE grant IBPG-0499-1.03/11.

References

[1] Brousell L. Five things you need to know about smart cities. 2012. Available at: http://bit.ly/TZRuu9.

[2] Hollands RG. Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? City, 2008;12(3): 303-320.

[3] Afonso RA, Brito KS, Nascimento C, Alvaro A, Garcia VC. Brazilian smart cities: using a maturity model to measure and compare inequality in cities. International Conference on Digital Government Research; 2015; Phoenix, Arizona, USA.

Address for correspondence

Ricardo Alexandre Afonso, Federal University of Pernambuco, (UFPE), Jornalista Anibal Fernandes, Ave,, 50.740-560 - Recife – PE; Phone: + 55 82 9184-8184., Contact: raa4@cin.ufpe.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-984

Factors Influencing Consent for Electronic Data Linkage in Urban Latinos

Suzanne Bakken[a,b], Sunmoo Yoon[a], Niurka Suero-Tejeda[a]

[a] *School of Nursing, Columbia University, New York, NY, USA*
[b] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*

Abstract

Within the context of patient participation in a Learning Health System, this study examined consent rates and factors associated with consent for linking survey data with electronic clinical data in a sample of 2,271 Latinos. Consent rate was 96.3%. Government insurance status and health literacy significantly influenced the odds of consent.

Keywords:

Learning Health System; consent; Latinos

Introduction

In the United States (US), as well as elsewhere, there is an increasing emphasis on the Learning Health System (LHS), which is enabled through a digital infrastructure. In its recommendations for the LHS, the Institute of Medicine highlighted the need for broader data capture and re-use[1]. A key component of the LHS is a fabric of trust (i.e. the willingness of individuals, and by extension, society, to contribute personal data and clinical experiences to the development of an LHS). However, it is difficult to build a strong fabric of trust among racial and ethnic minorities as reflected by low participation rates in research studies, low participation rates in biobanks, and limited use of information technologies for health-related purposes.

In previous research, we examined the sociodemographic factors and health factors that influenced consent for collection, storage and use of biospecimens in four contexts, with varying consent rates (53.2%-58.0%), in a sample of Latinos (n=2,290) who participated in the Washington Heights Inwood Informatics for Comparative Effectiveness Research (WICER) survey. Particpants were compensated with $10 for a saliva specimen and $15 for a dried blood spot. Government insurance and immigrant status increased the likelihood of consent in all contexts; higher health literacy increased the likelihood of consent for general use and long-term storage as compared for use by WICER team.

The purpose of this study was to address another decision made by individuals at the time of survey participation, which is relevant to the fabric of trust for the LHS, consent for linkage of survey data with electronic clinical data for research purposes. Two research questions were addressed: 1) Do consent rates for data linkage differ from those for biospecimen collection, use, and storage? 2) What sociodemographic and health factors are associated with the decision to link survey and electronic clinical data?

Materials and Methods

As part of the WICER survey in a primarily Latino immigrant community in New York City, a community health worker asked Latinos if they were willing to have their survey data linked with electronic clinical data. We examined factors associated with consent through logistic regression.

Results

In the sample of 2,271 Latinos, the consent rate for data linkage was 96.3%. This was significantly different than consent rates for all contexts of biospecimen use with Chi-square ranging from 78.6-92.8 ($p<.001$). Variables that met the criterion for entry into the logistic regression ($p<.2$ in bivariate analysis) were government insurance status, health status, hypertension, and health literacy. Gender, age, education, and immigrant status were not entered. Only two variables were significant: government insurance (OR=.332[.442-.775], p=.011) and health literacy (OR=1.35[1.09-1.67], p=.006).

Conclusion

Overall consent rates were high, suggesting a positive contribution to advancing the fabric of trust for a LHS. However, in contrast to our previous work on biospecimen consent, those with government insurance (Medicare or Medicaid) as compared to private insurance, a measure of socioeconomic status, were less likely to consent to linkage of survey data with electronic clinical data. Further work is needed to elucidate the role of financial incentives versus other factors in influencing Latinos to participate in LHS activities.

Acknowledgments

Funded by R01HS022961 (WICER4U) from AHRQ.

References

[1] *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*, Institute of Medicine of the National Academies, 2012.

Address for correspondence

Suzanne Bakken, sbh22@columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-985

Communication problems between end-users and technicians through a Help Desk in a Health Information System

Gabriela Garcia , Agustina Bertoia , Leonel Cameselle , Sonia Benitez , Diego Giunta , Analía Baum , Fernán Gonzalez Bernaldo de Quiros

Hospital Italiano de Buenos Aires, Argentina

Abstract

Communication between users and technicians is crucial for improving Help Desk performance. The objective of this study is to know and understand perceptions, and needs of users and help desk technicians. A qualitative study based on interviews was performed. The emergent topics were communication, workload and misperceptions between end-users and technicians. There are false perceptions between them that affect their interaction and work dynamics.

Keywords:

Qualitative research; perceptions; service support; Health information system

Introduction

End-users of Health Information Systems (HIS) must have adequate support and training to ensure proper HIS use. Adequate support requires human and technological resources capable of serving, managing and solving all of the possible problems, and good interpersonal communication skills between users and technicians is needed. Support and training are important components of HIS implementation but procedures, challenges and lessons learned are under-represented in the literature. A lack of support in complex scenarios like health care organizations may interrupt workflow causing dissatisfaction and impact in patient care. However, little is known about how end-users and technicians perceive their roles, their commitment and how they face communications challenges using a Help Desk (HD). The objective of this research is to understand end-users and technicians insight about their role in the support process.

Methods

Setting

In 1998 the Health Informatics Department (HID) of Hospital Italiano de Buenos Aires (HIBA) started the development and implementation of an HIS. The HID has 150 members from different disciplines. The computational infrastructure consists of 4,000 personal computers and 1,500 devices (printers, scanners, tablets) available for almost 8,000 users. Users can request support using an online HD. Each request generates a ticket that facilitates problem tracking. When the problem is solved, the user receives a survey by email to give feedback about the support.

Design

A Qualitative study based on in-depth interviews was conducted between March and September of 2014. Content analysis was performed using Grounded Theory and NVivo 10.

Results

Eighteen users (nurses, physicians, other professionals and administrative personnel) and fifteen HD technicians were interviewed. The emergent topics were: communication, workload (including perception of workload, priority of tasks, type of problem, type of end-user, feedback), and misperceptions between end-users and technicians. The most important are communication and misperceptions.

Technicians say they need to contact the user who made the request due to lack of information in the ticket description; highlighting the importance of verbal communication with the user. *Users* also believe it is better to communicate by telephone with technicians before requesting support because they doubt that it is well written or they feel there is a lack of information to explain the problem.

The *technicians* have the perception that users: (1) think their requests have high priority; (2) do not cooperate to be located easily; (3) are dissatisfied with the delay on ticket resolution; (4) do not value the technicians' work, because users do not give any feedback. On the other hand, *users* have the perception that the technicians do not try to communicate with users. However, users recognize that is difficult to be found because of the work dynamics, and understand that the delay in resolution is due to the HD workload. Finally, users admit that they do not value the technicians' work enough.

Conclusion

Users and technicians are the main actors in the HD process. To ensure good support of an HIS, it is necessary to implement communication strategies to remove barriers that create false perceptions.

Address for correspondence

Gabriela Garcia
Hospital Italiano de Buenos Aires
monica.garcia@hospitalitaliano.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-986

Speech therapy teleconsultations of a public telehealth service in a developing country

Aline Moreira Lucena[a], Maria Beatriz Alkmim[a], Vinícius Soares Garcia[a,b], Erica de Araújo Brandão Couto[b], Milena Soriano Marcolino[a,b]

[a]*Telehealth Center, University Hospital, Universidade Federal de Minas Gerais; Telehealth Network of Minas Gerais, Brazil*
[b]*Medical School, Universidade Federal de Minas Gerais, Brazil*

Abstract

Our aim is to assess speech therapy teleconsultations performed by the Telehealth Network of Minas Gerais, a public telehealth service that attends 722 cities in Brazil, to demonstrate the importance of the telehealth to support these professionals. In this observational retrospective study, consecutive speech therapy teleconsultations performed by the Telehealth Network of Minas Gerais, Brazil, from February 2011 to May 2014, were analyzed. Socio-demographic characteristics of the cities which requested teleconsultations were assessed, and teleconsultations were analyzed based on the type of query. Throughout the study, 259 valid speech therapy teleconsultations were performed. There were no significant differences in socio-demographic characteristics of municipalities that requested and did not request speech therapy teleconsultations. Speech therapists (65%), nurses (27%) and physicians (5%) requested the highest number of teleconsultations, mostly related to the area of language (47%), oral motor skills (29%), voice (20%), audiology (18%) and dysphagia (10%). In conclusion, teleconsultations demonstrated to be a potential tool for speech therapists working in remote areas.

Keywords:
Speech Therapy; Telemedicine; Public Health

Introduction

In Brazil, there is a high concentration of the healthcare professionals and resources in the largest cities. The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service that connects six university hospitals with primary care of 722 remote municipalities performing teleconsultations in several health specialties. The speech therapy teleconsultations began in TNMG since 2011. The objective of this study was to assess speech therapy teleconsultations performed by the Telehealth Network of Minas Gerais. A public telehealth service which attends 722 cities in Brazil, to demonstrate the importance of the telehealth to support these professionals.

Materials and Methods

This retrospective observational study assessed all consecutive teleconsultations performed by the TNMG, from February 2011 to May 2014. The teleconsultations were classified according to the professional who requested and the specialist who answered them. Socio-demographic characteristics of the cities which requested teleconsultations were also assessed.

Results

Throughout the study, 259 valid speech therapy teleconsultations were performed. Among the 668 cities attended by the TNMG, 81 requested at least one teleconsultation (12%). A comparative analysis of socio-demographic characteristics of users and non -users municipalities is shown on Table 1.

Table 1 – Socio-demographic characteristics of the cities

Variables	Users (n=81)	Non users (n=587)
Total population (x1,000 inhabitants)	6.40 (4.21-9.27)	6.724 (4.43-11.80)
HDI	0.66 (0.62-0.68)	0.66 (0.63-0.70)
Income per capita (USD)	231.31 (178.54-306.14)	252.15 (198.60-311.24)
% of poor	17.50(9.69-32.98)	17.53(9.22-26.98)

FHP: family health program, HDI: human development index. P-value was >0.05 for all variables.

Speech therapists (65%), nurses (27%) and physicians (5%) requested the highest number of teleconsultations. Regarding the type of query, 19% were educational, and 81% were related to the assistance of a particular patient. Of these, 35% were about diagnosis and 65% for discussion of procedures and therapeutic approaches. The questions were mostly related to the area of language (47%) and oral motor skills (29%), followed by voice (20%), audiology (18%), dysphagia (10%) and public health (3%). There was no significant relation between the professional category of the professional who requested the teleconsultation and the type of query.

Conclusion

Teleconsultations demonstrated to be a potential tool for speech therapists working in remote areas. The use of teleconsultations in speech therapy is still incipient, and the results indicate that there are no socio-demographic differences that justify the limited use of teleconsultation. We believe that the service must be more promoted to speech therapists that work in remote municipalities.

Address for correspondence

Aline Moreira Lucena: Universidade Federal Minas Gerais, Avenida Professor Alfredo Balena, 110. Santa Efigênia, BH. CEP 30130-100.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-987

Teleconsultations to Provide Support for Primary Care Practitioners and Improve Quality of care – the Experience of a Large Scale Telehealth Service in Brazil

Milena Soriano Marcolino[a,b], Julia Pereira Afonso dos Santos[a,b], Daniel Santos Neves[a], Maria Beatriz Moreira Alkmim[a]

[a]Telehealth Center, University Hospital, Universidade Federal de Minas Gerais; Teleheatlh Network of Minas Gerais, Brazil
[b]Medical School, Universidade Federal de Minas Gerais, Brazil

Abstract

This observational restrospective study was designed to assess teleconsultations performed by a public telehealth service in Brazil. A satisfaction survey was used to assess the impact on clinical practice. A total of 63,975 teleconsultations were performed, from April 2007 to November 2014. Family physicians (33%) and dermatologists (19%) answered most queries. From January to February 2014 (n=895), the most frequent queries were about etiology (30%) and pharmacological treatment (25%). The satisfaction survey in 2014 (n=571) showed that teleconsultations avoided patient referal in 78%. This study aims to shows the potential of telehealth to provide support to primary care practitioners in remote cities.

Keywords:

Telehealth, Public Health, Primary Healthcare.

Introduction

In Brazil, there is a high concentration of healthcare professionals and resources in the largest cities [1]. Telehealth was implemented to reduce this inequality in healthcare. The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service which connects six university hospitals with primary care units in remote cities. Since 2007, the network has been performing teleconsultations (second opinion) for a broad range of medical and non-medical specialties. The TNMG currently attends primary care practitioners in 722 cities [2].

The objective of this study is to assess the teleconsultations performed by the TNMG, in order to demonstrate the importance of telehealth to support these healthcare professionals, as well as promote continued education.

Methods

In this retrospective observational study, all teleconsultations performed between April 2007 and November 2014 were assessed and classified by both the requesting professional and the responding specialist. The teleconsultations were analyzed based on the type of query. A satisfaction survey was applied in the end of each teleconsultation to assess the impact on clinical practice.

Results

The study reviewed 63,975 teleconsultations that were performed. The healthcare professionals who requested the highest number of teleconsultations in the last 12 months were nurses (46%) and physicians (42%). The response time was under 24h for professionals on duty (family physicians, dermatologists, nurses, speech therapists, physiotherapists, dentists, psycologists) and under 48h for subspecialists. Family physicians (33%), dermatologists (19%), nurses (13%), obstetricians and gynecologists (10%), pediatricians (7%) and dentists (4%) answered the most queries. From January to February 2014 (n=895), 81% of the queries were related to patients' assistance and 19% were theoretical queries. The most frequent queries were about etiology (30%), pharmacologic treatment (25%), propedeutics (12%), non-pharmacologic treatment (12%). According to the ICD-10 classification, the majority of the queries were about skin and subcutaneous tissue diseases (16%), infectious and parasitic diseases (11%), digestive (8%), genitourinary (7%), cardiovascular (5%) and endocrine (5%) diseases. The satisfaction survey in 2014 (n=571) showed that teleconsultations avoided patient referal in 78% of cases, and 95% of the users were satisfied with the service.

Conclusion

This study shows the potential of telehealth to provide support and promote continued education for primary care practitioners in remote locations, as well as improving the access of the Brazilian population to specialized care.

References

[1] Conselho Federal de Medicina. Demografia Médica no Brasil. In: CREMESP. Cenários e indicadores de distribuição. 2013. p 256.

[2] Alkmim MB, Figueira RM, Marcolino MS, et al. Improving patient access to specialized health care: the Telehealth Network of Minas Gerais, Brazil. Bull World Health Organ. 2012;90(5):373-8

Address for correspondence

Milena S Marcolino: Avenida Professor Alfredo Balena, 190, room 246. CEP 30130-100. Email: milenamarc@gmail.com.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-988

Clinical Quality Control of a Large-Scale Teleconsultation Service

Maria Beatriz M. Alkmim[a], Milena Soriano Marcolino[a,b], Junia Xavier Maia[a,b], Cristiane G. Pessoa[a], Elaine Machado[a], Lidiane Sousa[a]

[a]*Telehealth Center, University Hospital, Universidade Federal de Minas Gerais; Teleheatlh Network of Minas Gerais, Brazil*
[b]*Medical School, Universidade Federal de Minas Gerais, Brazil*

Abstract

The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service in Brazil that assists 722 municipalities in the state of Minas Gerais. As a large-scale teleconsultation service, it was important to implement clinical quality control to guarantee the quality of the service. Our aim is to describe the audit of the teleconsultation responses performed by TNMG. A random sample was selected from teleconsultations performed by the specialists from the TNMG between January and February 2014. The responses were evaluated regarding size, objectivity, quality, ethics, courtesy and grammar. A total of 640 teleconsultation responses were assessed, and the mean scores were ≥2.45. Objectivity and quality had the lowest scores in the different specialities. The methodology was useful for evaluating the teleconsultation service and for identifying the areas to improve.

Keywords:

Telemedicine, Teleconsultation, Public Health, Quality Control.

Introduction

Telehealth has been growing rapidly worldwide. Teleconsultation is an important tool to support healthcare professionals in remote municipalities. It has high potential to improve access to healthcare and improve quality of care at reduced costs. The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service in Brazil that assists 722 municipalities in the state of Minas Gerais. As the service expands it is important to perform periodic audits to guarantee its quality. The objective of this study is to describe the audit of the responses of teleconsultations performed by the TNMG.

Materials and Methods

A random sample was selected from teleconsultations performed by the specialists from the TNMG between January and February 2014. The responses were scored from 1 to 3. Different weights were attributed to each variable according to its impact on the quality of the teleconsultation. Size was weighted by a factor of 0.5, objectivity and quality by 0.3, ethics by 0.1, and courtesy and grammar by 0.05. All medical and non-medical specialities and subspecialities were assessed. For the purpose of this study only the results of the professionals who work on duty will be shown.

Results

A total of 640 teleconsultation responses were assessed, of which 76% were medical and 24% non-medical. Of the non-medical responses, 65% were nursing, 12% dentistry, 8% physiotherapy, 6% pharmacy, 5% psychology, 3% nutrition and 2% speech therapy. Table 1 shows the objectivity, quality, ethics and courtesy scores for medical and non-medical teleconsultation responses. Scores for size ranged from 2.67 to 3.00 (medical) and 2.74 to 3.00 (non-medical), and for grammar ranged from 2.79 to 3.00 (medical) and 2.97 to 3.00 (non-medical). The final score ranged from 2.74 to 2.98 for medical and from 2.75 to 3.00 for non-medical specialities.

Table 1 – Scores for medical and non-medical teleconsultations responses

	Objectivity	**Quality**	**Ethics**	**Courtesy**
Medical	2.90	2.71	2.91	2.99
Dermatology	2.99	2.92	2.75	3.00
Family medicine	3.00	2.79	3.00	3.00
Gynecology	3.00	2.88	2.90	2.98
Internal Medicine	2.71	2.58	2.96	3.00
Pediatrics	2.58	2.45	2.93	3.00
Surgery	2.89	2.44	3.00	3.00
Non-medical	2.71	2.68	2.97	3.00
Nursing	2.61	2.64	2.97	3.00
Dentistry	2.84	2.79	3.00	3.00
Physiotherapy	2.85	2.54	3.00	3.00
Pharmacy	2.89	2.78	3.00	3.00
Psychology	3.00	3.00	3.00	3.00
Nutrition	3.00	2.75	3.00	3.00
Speech therapy	3.00	3.00	3.00	3.00

Conclusion

The innovative methodology used to audit TNMG teleconsultations was useful for identifying aspects to improve with corrective actions to guarantee the quality of the service.

Address for correspondence

Maria Beatriz Alkmim. Avenida Professor Alfredo Balena, 110, 1º Andar. CEP 30130-100. E-mail: balkmim@yahoo.com.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-989

Audit of primary care electrocardiograms sent as emergency to a telehealth service - the Telehealth Network of Minas Gerais, Brazil

Milena S. Marcolino[a,b], Bárbara C.Carvalho[a], Aline M. Lucena[a], Ana Luiza N. França[a,b], Cristiane G. Pessoa[a], Daniel S. Neves[a], Maria Beatriz M. Alkmim[a]

[a]*Telehealth Center, University Hospital, Universidade Federal de Minas Gerais; Teleheatlh Network of Minas Gerais, Brasil*
[b]*Medical School, Universidade Federal de Minas Gerais, Brazil*

Abstract

The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service in Brazil that has performed electrocardiogram (ECG) analysis since 2005. From February to March 2014, 28% of ECGs were classified as "emergency" by the primary care tele-health sites. This quasi-experimental study aimed to investigate the reasons behind the high number of emergency ECGs being sent in, the implementation of corrective actions, and an assessment of the impact of these actions. In the 1st phase, primary care units that sent >70% of ECGs as emergency from February to March 2014 were selected. The 2nd phase consisted of the intervention. In the 3rd phase, the proportion of ECGs sent as an emergency during the 1st and 2nd months post intervention were assessed. Of the 63 cities selected during the 1st phase, 50% of the practitioners did not know the proper definition of emergency. After the intervention, 67% of the cities had a significant reduction in the proportion of ECGs sent as an emergency during the 1st month, and 17% had a significant reduction during the 2nd month.

Keywords:

Telemedicine, Eletrocardiography, Primary Healthcare.

Introduction

The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service that assists 730 municipalities in the State of Minas Gerais, Brazil. From February to March 2014, 28% of ECG exams performed by the primary care telehealth sites were classified as "emergency". Our hypothesis was that a great number of ECGs were incorrectly classified as "emergency". This is potentially harmful, as it delays the assessment of the true emergency ECGs. The objective of this study was to investigate the reasons for the primary care telehealth sites sending a higher number of emergency ECGs than expected, and to implement corrective actions in order to reduce the proportion of ECGs incorrectly classified as emergency.

Materials and Methods

In the first phase of this quasi-experimental study, primary care units that sent >70% of ECGs as emergency from February to March 2014 were selected to participate. The 2nd phase consisted of open question telephone calls to the 63 telehealth sites selected to investigate the criteria adopted to classify a case as an emergency. Primary care practitioners were informed about the criteria to classify an ECG as emergency using Brazil's Federal Council of Medicine. An e-mail with reinforcement information was also sent to each practitioner. In the 3rd phase, the proportion of ECGs sent as an emergency during the 1st and 2nd months after the intervention were assessed. Chi-square tests were performed to compare baseline, 1st and 2nd months.

Results

Of the 63 telehealth sites, nurses and nursing technicians were responsible for 92% of the ECGs performed. 50% of the practitioners did not know the proper definition of an emergency. All exams were sent as emergency by 16% of primary care professionals.

After the intervention, 67% of the sites had a significant reduction ($p < 0,05$) in the proportion of ECGs sent as an emergency during the 1st month post intervention, and 17% had a significant reduction during the 2nd month (Figure 1).

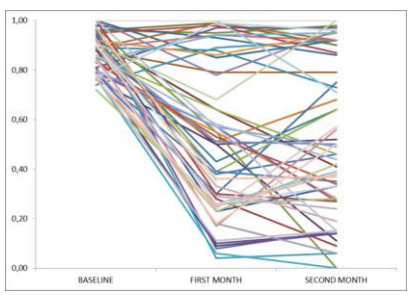


Figure 1- Proportion of ECGs classified as emergency at baseline, 1 month and second month post intervention

Conclusion

This study demonstrated that many medical professionals responsible for ECG exams at the primary care units in remote cities did not know the correct criteria definition of emergency, and that interventions as simple as a telephone call and email may be effective in reducing the proportion of ECGs incorrectly classified as an emergency.

Address for correspondence

Milena S Marcolino: Avenida Professor Alfredo Balena, 190, room 246. CEP 30130-100. Email: milenamarc@gmail.com.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-990

Referring Quality Assessment of Primary Health Care for Endocrinology in Rio Grande do Sul, Brazil

Gabriela Monteiro Grendene[a], Átila Szczecinski Rodrigues[a], Natan Katz[a], Erno Harzheim[a]

[a] *TelessaúdeRS - Programa de Pós Graduação em Epidemiologia, UFRGS, Brazil*

Abstract

This paper presents results of an assessment of the quality research of endocrinology referrals in the public health system in the state of Rio Grande do Sul. From the analysis of 4,458 requests for endocrinology referrals, it was found that 15% of referrals had insufficient information for evaluation and 71% showed no clinical justification for authorization of referencing. The partial results of the study indicated that the lack of information makes it impossible to clinically regulate these requests. The use of referencing protocols associated with telemedicine tools can assist doctors in primary health care in the clinical management and make access to specialized services more equitable and timely.

Keywords:
Endocrinology; Telemedicine; Telehealth.

Introduction

The project Regulasus is a partnership between the core TelessaúdeRS / UFRGS, of the Federal University of Rio Grande do Sul (UFRGS), and the State Government of Rio Grande do Sul, that qualifies and reduces medical specialty referrals [1]. One of the biggest barriers in the internal medicine regulatory process is the lack of clinical information, which makes it difficult to evaluate the need for consultation and priority of access to health care in various specialties [2, 3]. This study aims to evaluate the quality of clinical information of endocrinology referrals. It is based on referencing protocols approved by the State Complex Governor of Rio Grande do Sul.

Materials and Methods

TelessaúdeRS / UFRGS researchers have developed referral protocols for the six most common medical conditions in the area of endocrinology (diabetes mellitus, hypothyroidism, hyperthyroidism, thyroid nodules, multinodular goiter and obesity). Based on these protocols, a medical regulator reviewed the requests, between November 2013 and December 2014, using the following classification:

- Reference with clinical justification for consultation with an endocrinologist (authorized request);
- reference without clinical justification for consultation with an endocrinologist (TelessaúdeRS consulting);
- reference without sufficient clinical information to regulation (pending for lack of information).

Results

There was a total of 4,363 regulated requests for Endocrinology. Of this total, only 621 (14%) were considered appropriate and sent for consultation, 3081 cases (71%) were referred for teleconsulting with TelessaúdeRS/UFRGS team, and 634 requests (15%) were returned to municipalities.

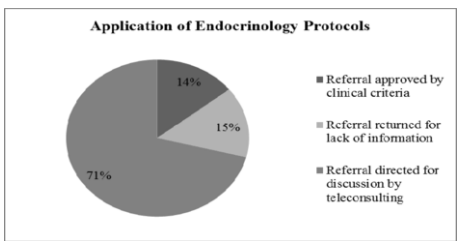


Figure 1 - Application of Endocrinology Protocols

Conclusion

The partial results of this study indicate that the lack of information makes regulation of clinical referrals difficult. Patients whose clinical management can be performed in Primary Health Care make the demand much greater than the supply, and hinder access to other patients for consultation in specialized services. The use of referral protocols can help primary health doctors by making access to care faster and more equitable.

References

[1] Rio Grande do Sul. Resolução n. 170/14- CIB, de 7 de abril de 2014. Aprova os encaminhamentos para o projeto intervenção das teleconsultorias na regulação das consultas ambulatoriais na especialidade Nefrologia. Portuguese.

[2] Almeida PF, Gérvas J, Freire JM and Giovanella L. Estratégias de integração entre atenção primária à saúde e atenção especializada: paralelos entre Brasil e Espanha. Saúde debate. 2013: 37(98): 400-415.

[3] Ministério da Saúde (Br). Protocolos de encaminhamento da atenção básica para a atenção especializada: Endocrinologia e Nefrologia. Brasília: Ministério da Saúde; 2015.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-991

2,000,000 Electrocardiograms by Distance: An Outstanding Achievement for Telehealth in Brazil

Milena Soriano Marcolino[a,b], Maria Beatriz Moreira Alkmim[a], Leonardo Bonisson[a], Renato Minelli Figueira[a], Antonio Luiz Ribeiro[a,b]

[a]*Telehealth Center, University Hospital, Universidade Federal de Minas Gerais; Teleheatlh Network of Minas Gerais, Brazil*
[b]*Medical School, Universidade Federal de Minas Gerais, Brazil*

Abstract

Our aim is to describe the evolution of the telediagnostic service of the Telehealth Network of Minas Gerais (TNMG), a public telehealth service in Brazil. It started in 2006 with 82 cities, restricted to electrocardiography analysis. Currently it extends to 772 cities - performing also Holter, ambulatory blood pressure monitoring and retinography analysis - and 48 ambulances in the north of the state, as part of a myocardial infarction system of care. Using low-cost equipment and simple technology, TNMG has employed various strategies to increase telehealth use. The number of ECGs performed by TNMG has progressively increased. It was expected to achieve 2 million in February 2015. The utilization rates were around 90%. It proved to be economically sound, promoting savings of 45M USD for an investment of 10.2M USD. It is currently a regular health service in the state, integrated into the healthcare system. In conclusion, the telehealth model developed in Minas Gerais produced good clinical and economical results.

Keywords:
Telemedicine; Telecardiology; Public Health.

Introduction

Minas Gerais is a Brazilian state with important social, cultural, economic, infrastructural and geographical contrasts. The Telehealth Network of Minas Gerais (TNMG) is a public telehealth service formed by 6 public universities. It performs teleconsultation and telediagnostic services, to help improve healthcare for the population who live in distant and poor regions. The objective of this study is to describe the evolution of the telediagnostic service provided by the TNMG.

Materials and Methods

This is a descriptive study. The TNMG started in 2006 in primary care services of 82 cities. It was restricted to electrocardiography analysis in primary care. Using low-cost equipment and simple technology, TNMG has employed various strategies to improve/increase telehealth use. Financial support is provided by federal, state and municipal government. The service keeps an online technical support for the villages. The utilization rates are monitored by a specialized team. Progressive expansions were performed to other cities and other levels of care. New applications were developed and validated, to increase the exams performed - such as Holter, ambulatory blood pressure monitoring and retinography analysis. A new tele-electrocardiology application to use in Androids was also developed, to increase usability in ambulances. All applications are in Java, interoperable, and data is compacted into a database (PostgreSQL). Technology and maintenance methodology are constantly evaluated and improved. Periodic audit is implemented to assure quality of the ECG analysis.

Results

The TNMG currently extends to 772 of the 853 cities of the state, including 4 secondary care units, 7 emergency care units in Belo Horizonte, the state's capital, and 48 ambulances in the north of the state, as part of a myocardial infarction system of care. The number of ECGs performedwere expected to reach 2 million in February 2015. During the whole period, the monthly ECG service utilization rates were around 90%. This showed that the service was well accepted by practitioners and incorporated to the health system of the municipalities. Therefore, the telecardiology is nowadays a regular health service in the state, integrated into the healthcare system.

The service proved to be economically sound, promoting savings of 45M USD for an investment of 10.2M USD; an outstanding achievement for telehealth in Brazil (return of investment = 4.4 and benefit-cost ratio = 5.3).

The tele-elctrocardiography in the ambulances and the organization of the services in the north of the state are expected to decrease myocardial infarction mortality.

Conclusion

The telehealth model developed in Minas Gerais, Brazil, has produced good clinical and economical results. As a consequence, it is now a regular health service in the State, covering 722 of the 853 municipalities and integrated into the healthcare system. The model and technology characteristics permit replication in other parts of the world.

Address for correspondence

Milena S Marcolino: Avenida Professor Alfredo Balena, 190, room 246. CEP 30130-100. Email: milenamarc@gmail.com.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-992

Care Professionals' Perceived Usefulness Of A Rehabilitation Ehealth Service In Stroke Care

Nadia Davoody[a], Maria Hägglund[a]

[a] *Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Sweden*

Abstract

Despite many attempts to support stroke patients, there is still room for improvement. The aim of this study is to gain insight into care professionals' perceived usefulness of an online care and rehabilitation planning tool. A functional prototype was developed and presented to a neurology team in a primary care centre in Stockholm. Three focus group meetings were conducted. The data were analysed based on the unified theory of acceptance and use of technology. The results indicate that the care professionals were positive towards the tool and described potential usefulness such as ease of understanding the rehabilitation process and support for collaboration among care providers and also cooperation between the patient and the team. They also identified challenges such as time limitation in daily care.

Keywords:
Stroke; eHealth; Rehabilitation; Unified Theory of Acceptance and Use of Technology.

Introduction

Stroke rehabilitation is a cyclical process of identifying patients' problems, defining goals and activities and assessing patients' progress towards the agreed goals [1]. Post-discharge stroke patients in Stockholm County Council receive home-based rehabilitation from neurology teams up to one year. The teams establish a paper-based rehabilitation plan together with patients and possibly their next-of-kin. We designed a web-based prototype which provides a care and rehabilitation plan that includes *problems, goals, activities,* and *outcomes.* Other features are a *calendar, reminders, list of disabilities, care providers, assistive tools* etc. Patients can identify problems, set goals, and plan activities together with the neurology team or independently. As using the tool will affect the care professionals' work routines and job performance, we in this study explore their perceived usefulness of the prototype.

Materials and Methods

Three focus group meetings with members of a neuro team (speech therapist, occupational therapist, counsellor, and physiotherapist) were performed. A short description of the study was given in the beginning of each meeting, and the prototype was introduced. Each focus group meeting lasted about one hour. The meetings were audio recorded, transcribed verbatim and analysed deductively based on the Unified Theory of Acceptance and Use of Technology (UTAUT) [2] as it has been reported to explain 70% of the variance in intention to use information technology.

Results

The participants believed that the tool would be useful in their work and it would support the cooperation between the patient and the team; however, they believed that they would not be able to use it themselves due to time limitation (**Performance expectancy**). "I just think that we all are working MI [motivational interviewing] based today and this [the tool] gives the patient control of his/her own rehabilitation and we have the opportunity to work even more on MI…" (*Occupational therapist*) "If we want to work with this, there will not be much time over for training, so we will get stuck with it and with the computer." (*Physio therapist*)

Participants mentioned that the connection between problems, goals and activities would lead to a better understanding of the rehabilitation process (**Effort expectancy**) "It is good to connect them, then you understand why you do the exercises. It can sometimes affect the motivation." (*Speech therapist*) Participants also believed that the information content in the tool is imperative for patients to understand their journey.

Social influence was not identified as an important factor in this study.

Participants wished the program was available through mobile devices and integrated with the patient record system (**Facilitating conditions**). "Mobile applications would be great. It is something that the patient always has with him/her…" (*Occupational therapist*). Another important facilitating condition was the availability of technical support, both for healthcare professionals and for patients.

Conclusion

The care professionals were cautiously positive towards the tool stressing their limited time with the patients. Their acceptance was mainly dependent on the usefulness of the tool for team work and relationship between care professionals and the patient. Technical support played an important role in their decision about the acceptance of the tool.

References

[1] Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. The Lancet. 2011;377(9778):1693-702.

[2] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS quarterly. 2003:425-78.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-993

Routine Health Information Systems in South Africa - Opportunities for Improvement

Edward Nicol[a,b], Lyn A. Hanmer[a]

[a] *Burden of Disease Research Unit, South African Medical Research Council, South Africa*
[b] *Division of Community Health, Stellenbosch University, South Africa*

Abstract

A recurring theme in published studies is the need for the appropriate human and other resources to support routine health information system (RHIS) implementation. While training in the use of specific RHIS and the availability of the required resources for implementation are essential for all users, other factors such as managers' understanding of the role of RHIS in supporting health services; the ability to interpret RHIS data; and a focus on data quality are further requirements for effective RHIS implementation.

Keywords:
Data quality, Routine health information system (RHIS), Health information system (HIS), Human resources, Competence, Health system strengthening

Introduction

Two reviews of progress made towards strengthening South African health information systems (HIS) in the context of current health sector policy initiatives identified multiple challenges which have persisted over the years due to factors including managers' inability to translate key health regulations and policies into practice; inadequate human resources; insufficient capacity of health information personnel; inadequate training; lack of resources such as registers, computers, and printers; limited HIS development planning; and lack of established HIS career paths and accredited training programmes[1][2].

Methods

Four studies on the evaluation of South African HIS by staff of the South African Medical Research Council published between 2010 and 2013 [3-6] were reviewed and the results synthesised to identify opportunities for strengthening existing RHIS performance and for improving future RHIS implementations. The foci of the respective studies were the effective use of computerised hospital information systems (CHIS) [4], RHIS resources [3], data quality challenges [5], and factors affecting data quality, such as RHIS competence and knowledge of RHIS rationale [6].

Results

The following key recommendations are made, based on the results of the review:

- Continue and strengthen efforts to ensure the effective use of existing RHIS, including the District health information system (DHIS) and other RHIS in use at patient, facility, district and provincial level.
- Strengthen management commitment and support, including appropriate resource allocation for RHIS [3][4].
- Focus on data quality, including allocation of required personnel and other resources to enable active monitoring of RHIS use and RHIS reports [4][5][6].
- Allocate, train and support management and end users in the use of RHIS. Multiple approaches and content are required. Include interpretation and analysis of data from the RHIS in management training [4][6].

Conclusion

This review has documented challenges of RHIS implementation in South Africa, and has identified opportunities for improving effectiveness of existing and future implementations.

Acknowledgments

The study was supported by the South African Medical Research Council.

References

[1] English R, Masilela T, Barron P, Schönfeldt, A. Health information systems in South Africa. In: Padarath A, English R, editors. South African Health Review. Durban: Health Systems Trust; 2011. p. 81-90.

[2] Loveday M, Smith J, Monticelli F. A comprehensive baseline audit of the South African national health information system. Durban: Health Systems Trust; 2006.

[3] John Snow Inc, South African National Department of Health. District Health Information System: Rapid Information Needs Assessment. Pretoria: John Snow Inc.; 2012. Technical Report.

[4] Hanmer, LA, Isaacs S, Roode JD (2010). Factors associated with hospital information system success: Results of a survey in South Africa. In: Safran C, Reti S, Marin HF, editors. Studies in Health Technology and Informatics, Vol. 160: MEDINFO 2010. p347 - 351. September 2010.

[5] Nicol E, Bradshaw D. Maternal, newborn and child survival: data challenges. In: Fonn S, Padarath A, editors. South African Health Review. Durban: Health Systems Trust; 2010. p. 73-8.

[6] Nicol E, Bradshaw D, Phillips T, Dudley L. Human Factors Affecting the Quality of Routinely Collected Data in South Africa. In: Lehmann CU, Ammenwerth E, Nøhr C, editors. Studies in Health Technology and Informatics, Vol. 192: MEDINFO 2013. p788 - 792. August 2013.

Address for correspondence

Edward Nicol, Email: Edward.Nicol@mrc.ac.za

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-994

Perceived Reasons for High and Low Quality Observational HIV Research Data

Stephany N. Duda[a], Catherine C. McGowan[b], Cynthia S. Gadd[a]

[a] *Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA*
[b] *Department of Medicine, Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, TN, USA*

Abstract

Audits of data quality in a Latin America HIV research network revealed that study sites collected weight measurements, laboratory results, and medication data of inconsistent quality. We surveyed site personnel about perceived drivers of their high or low quality data. Most sites reported their research teams contained no data specialists and that missing data stemmed primarily from incomplete patient assessments at the point of care rather than inconsistent data recording. The root causes of data errors resulted from limited clinic resources (e.g., broken scales, limited record storage space), workflow complications, or the indifference of external participants towards research activities. Understanding these factors supports targeted quality improvement processes.

Keywords:

Data Collection, Error Sources, Data Quality

Introduction

The Caribbean, Central, and South America Network for HIV research (CCASAnet), a network of HIV clinics in Latin America, conducts on-site reviews of routine clinical care data submitted for research by study sites. During a data audit, data coordinating center personnel check data submitted for research against the contents of paper clinical records (source documents), flagging data mismatches, missing data, and data with no documented origin. This source document verification helps quantify data quality to ensure meaningful research findings [1]. The audits revealed many discrepancies in CCASAnet sites' data. We investigated the causes of data quality differences across sites through a data quality survey.

Materials and Methods

We developed a survey that asked site personnel about the presumed causes of their high or low quality data. We focused on 3 high-error categories: weights, lab values, and medication history. The survey was approved by the Vanderbilt IRB, implemented in REDCap (projectredcap.org), and distributed by coordinators at each site. Survey responses were grouped by site, but anonymous within sites. The project ran March-June 2009. Two researchers independently selected recurrent themes through iterative, bottom-up analysis. Conclusions were harmonized through group discussion. A third researcher was appointed to resolve any conflicting interpretations.

Results

We received 18 survey responses, representing 6/7 CCASAnet sites. Respondents reported that clinicians conducted data abstraction rather than data personnel, and teams included on average 3-5 people. Reported reasons for missing *weight data* included broken scales and providers' perception that weight measurements were unimportant for clinical care. Two sites with high quality weight data described reducing missingness via multiple data capture opportunities

Auditors often could not find source documentation for th*e laboratory data* sent for research. Survey responses described oversized laboratory printouts that didn't fit into medical record folders, a lack of shelf space in the record room, and the practice of giving lab reports to patients as reasons for missing paper documents. The one site with high quality lab data received direct data exports from the laboratory system.

Sites with low quality *medication data* agreed that brief drug regimens were easily overlooked during data abstraction and delays between when a physician prescribes treatment, the government approves the treatment, and the pharmacy dispenses the drugs resulted in multiple, conflicting dates recorded in the patient chart. Collaborative work with local clinic pharmacies was reported as a reason for higher quality data, and sites benefitted from having access to pharmacy records as a secondary source of information.

Discussion

The responses of survey participants revealed that higher quality research data was the result of complementary processes (e.g., parallel laboratory or pharmacy workflow) or automated data imports, which reflects well-documented quality improvements when reducing manual transcription [2]. Most on-site data preparation roles were filled by clinicians although trained data clerks may conduct accurate and more cost-effective abstraction of basic data [3]. Respondent feedback informed potential approaches to reduce errors in HIV research data abstracted from patient care records.

Acknowledgments

The authors thank Daniel R. Masys, MD, and the teams at the CCASAnet sites. This work was supported by NIH/NIAID U01AI069923 and NCATS UL1TR000445.

References

[1] Winchell T. The clinical investigator site audit – process-driven good clinical practice. Qual Assur J 2007;11:138–42. doi:10.1002/qaj.412.
[2] Arts DGT, et al. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. JAMIA 2002;9:600–11.
[3] Mi MY, et al. Reliability of medical record abstraction by non-physicians for orthopedic research. BMC Musculoskelet Disord 2013;14:181.

Address for correspondence

Stephany Duda, stephany.duda@vanderbilt.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-995

Telehealth in Brazil: Contemporary Tool for Access to Health

Thais Vieira Esteves[a] and Sérgio Pacheco de Oliveira[a]

[a] *National School of Public Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil.*

Abstract

This work raises questions about telehealth in Brazil, especially the areas of diagnosis, treatment and monitoring at distance, emphasizing its importance for the improvement of health conditions. It was based on a literature review. Three successful experiences were selected as examples: The Minas Gerais Project Telecardio, the São Paulo University Teleaudiology and Telerehabilitiation of UnB. Despite the increase of telehealth experiences in Brazil, much remains to be done in regard to diagnosis, treatment and follow-up, with potential positive effects on health.

Keywords:
Telehealth; Diagnosis; Treatment ; Monitoring.

Introduction

Internet technology has produced impacts on society and therefore, in health. The telehealth field, where communication for healthcare technologies is incorporated, is an important example for our country, due its geographical size and inequalities of distribution of health care resources. This study is a literature review that aims to show the level of development of telehealth in Brazil, focusing on experiences related to diagnosis, treatment and monitoring.

Materials and Methods

A literature search was conducted using the PubMed, SciELO and Scopus databases. The search terms were "internet" and "health," using the free form, selected for all levels, from 2010 to 2014. We found more than 200 articles and by reading the titles, we selected 73 texts for a full-content review. We observed just a few experiences in diagnosis, treatment and follow-up in Brazil. We selected three successful experiences to demonstrate the advances and challenges in the area.

Results

In the late 80s, telehealth began in Brazil, mainly at universities and research centers, increasing access to health care for residents of remote areas and reducing inequalities of health services. In 2010, the Ministry of Health created the Telehealth Program Brazil, which was redefined and expanded in 2011 with the goal of supporting the consolidation of the Health Care Networks in SUS. Although we have in Brazil, just a few experiences of diagnosis, treatment and follow-up, this paper selected three successful experiences as examples of improvement to the Brazilian heath system.

The Telecardio Project implemented a low cost telecardiology system in the small cities of Minas Gerais state. This enabled diagnosis in cardiology support for primary care physicians and improved referrals of complex cases, in addition to assisting the permanent education of municipalities' professionals. Today it is a permanent government program in the state of Minas Gerais.

The experience in Teleaudiology from São Paulo University was characterized by adapting individual hearing aids at distance. It was observed that in addition to adaptation, it allowed the training and preparation of professionals from a remote center using a specific hearing amplification software.

The Brasilia's University Telerehabilitation investigated the use of Internet for people with spinal cord injury. It proposed the improvement of care in rehabilitation services routines using the Internet as a tool for the well-being of patients.

Conclusion

Despite the growth of telehealth practices in Brazil, a lot still needs to be done. Telehealth is an important tool to support health actions, facilitating access to early diagnosis and decreasing the time to start of treatment, so the prognosis of the disease tends to be improved and the health cost decreased. On the other hand, Brazilian telehealth requires more infrastructure investment and stimulus as shown by the success cases, for early diagnosis, treatment and follow-up at a distance.

References

[1] Nepal S, Li J, Jang-Jaccard J and Alem L. Case study: a Framework for Telehealth Program Evaluation. Telemedicine and e-Health 2014: 20 (4): 393-404.

[2] Ribeiro ALP, Alkmim MB, Cardoso CS, Carvalho GGR, Caiaffa WT, Andrade MV, Cunha DF, Antunes AP, Resende AGAR and Resende ES. Implantación de um Sistema de Telecardiología em Minas Gerais: Proyecto Minas Telecardio. Arq Bras Cardiol 2010: 95(1): 70-78.

[3] Penteado SP, Ramos SL, Battistella LR, Marone SAM and Bento RF. Remote hearing aid fitting: Tele-audiology in the context of Brazilian Public Policy. Int Arch Otorhinolaryngol 2012: 16(3): 371-381.

[4] Rodrigues MPC and Araujo TCCF. Internet as a Support for People with Spinal Cord Injuries: Usage Patterns and Rehabilitation. Paidéia 2012: 22(53), 413-421.

Escola Nacional de Saúde Pública. Rua Leopoldo Bulhões, 1480. Manguinhos, Rio de Janeiro, Brazil. CEP: 21041-210.

E-mail: thaisesteves@fiocruz.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-996

Stigma and On-line Health Information Seeking of U.S. South Asian Cancer Survivors

Sheba M. George, PhD [a, b] and Marjorie Kagawa Singer, PhD, MA, MN, RN, FAAN [a, c]

[a] Department of Community Health Sciences, UCLA Fielding School of Public Health; [b] Center for Biomedical Informatics, Charles R. Drew University of Medicine and Science, Los Angeles, CA.; [c] UCLA Asian American Studies Center, Los Angeles, CA.

Abstract

The internet has replaced physicians as primary health information source for cancer-survivors.It is important to uncover barriers/facilitators to cancer information seeking, particularly on-line.Asian Americans are the fastest growing U.S racial/ethnic minority, 2) cancer is the leading cause of r death and 3) cancer knowledge is low among them and little research is done on their cancer information seeking strategies. This study aims to examine qualitatively cancer information-seeking patterns of the Asian American group, South Asians, using in-depth interview methods. Family members and social networks are highly engaged in providing informational support to South Asian cancer survivors. such collaborative information seeking is limited by stigma related to cancer and must be taken into consideration when developing culturally appropriate cancer health information seeking interventions in such communities.

Keywords:

Health information seeking; cancer survivors; Asian American

Introduction

As cancer survivorship has become acknowledged as an important stage in the cancer continuum of care, health information seeking by cancer survivors is a major under-addressed area of need. The widespread access to readily available on-line health information has been characterized as having the potential to level the playing field for all patients, particularly for underserved racial and ethnic minority populations experiencing health disparities. Yet, Asian American survivors and familiesreceive limited guidance on health information seeking.[1]

Materials and Methods

A 90-minutes to 2 hours in- in-depth interviews were conducted with a sample of 40 participants, including 20 dyads of 20 cancer survivors, varying by level of education and 20 information-support providing family members/friends. South Asians included those having ethnic origins in the Indian subcontinent, including India, Pakistan, Bangladesh and Sri Lanka. Questions addressed a range of domains including expectations, experiences and needs that shape cancer health information seeking.

Results

Family members and social networks are highly engaged in providing informational support to South Asian cancer survivors, as is the case for other ethnic and minority populations.[2] In South Asian communities, differences in norms about patient autonomy and cultural/family values, linguistic challenges and health literacy may result in greater reliance of cancer patients on family members and social networks, such that they function as a unit, and in some cases, family members may conduct all information seeking for the cancer survivors. Such family-based information seeking behavior and patterns of communication go against assumptions about patient autonomy in the U.S. health care.

We found that their collaborative information seeking was limited by stigma or negative reactions to cancer in the community, which led to selective sharing of cancer with family and community. Participants characterized cancer stigma as: 1) judgment from God; 2) a marker of hereditary genetic failing and 3) a marker of lost social status, such as mastectomies, where women can be seen as not fully women.

Conclusions

To better understand health information seeking of Asian Americans, particularly recent immigrants, models of culturally collaborative or surrogate information seeking are needed compared to existing models assuming autonomous individual information seeking. For many South Asian cancer survivors, the fabric of life is intertwined with that of family/friends due to cultural, logistical, and linguistic dependency. Such a conceptualization would allow interventions to adapt health information resources/channels to relevantly meet the needs of these sub-populations. The explicit stigma against cancer points to the value of understanding the sources of such stigma, its impact on cancer health information seeking and developing appropriate strategies to deliver cancer information resources to such populations, particularly those with language/health literacy deficits.

Acknowledgments

This work was supported by the National Cancer Institute, NIH (grant 3U54CA153499-04S1)

References

[1] Chen Jr MS. Cancer health disparities among Asian Americans. *Cancer.* 2005;104(S12):2895-2902.

[2] Watts T., Merrell J., Murphy F. & Williams A. (2004) *Journal of Advanced Nursing* 47(5), 526–535 Breast health information needs of women from minority ethnic groups

Address for correspondence

Sheba George can be contacted at shebageorge@ucla.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-997

End-User Experiences and Expectations Regarding Data Registration and Reuse Before the Implementation of a (New) Electronic Health Record: A Case Study in Two University Hospitals

Erik Joukes[a], Nicolette de Keizer[a], Ameen Abu-Hanna[a], Martine de Bruijne[b], Ronald Cornet[a,c]

[a] *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, the Netherlands*
[b] *Department of Public and Occupational Health, EMGO+ Institute for Health and Care Research, VU University Medical Centre, Amsterdam, the Netherlands*
[c] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract

Patient data stored in Electronic Health Records (EHRs) are used during care provision but are also potentially usefully reused for other purposes. Data (re)use requires good data quality, which necessitates efforts by healthcare professionals for proper data registration. However, their commitment depends on their perception of the reuse benefits. We developed a questionnaire to investigate the perception and expectations of end-users on data registration and reuse in two university hospitals starting a joint EHR implementation. Especially personnel in direct patient care reports to spend much time (40%) on data registration and this group is not willing to spend more time with the new EHR. Additionally, approximately one third of the personnel did not yet have a clear view on future developments regarding data registration and reuse. We found only small differences between hospitals.

Keywords:
Medical Record Systems; Health Personnel; Data Registration; Electronic Health Records; Implementation; Data Reuse.

Introduction

Reuse of patient data stored in an Electronic Health Record (EHR) relies on data of sufficient quality and structure. The quality is influenced by the EHR and the personnel's attitude towards the importance and benefits of data registration but also the amount of time spent on registration. The implementation of a new EHR offers an opportunity to improve data quality. We investigated 1) how much time the personnel thought to spend on registering data now and how much they were willing to invest when working with a new EHR, and whether this differs between a hospital currently working with a legacy EHR versus one currently registering on paper, 2) how much faith the personnel has in the (future) registration and reuse of patient data.

Methods

Based on focus groups and a literature search, we developed an online questionnaire to measure the expectations of the (future) EHR end-users on data registration and reuse before the implementation of a new EHR. We invited for the questionnaire approximately 10,000 (future) EHR end-users of two university hospitals in the Netherlands (one with a legacy EHR, the other registering on paper) currently in the process of jointly implementing a new EHR. The majority of the personnel was not familiar with the functionality and effects of the new EHR.

Results

Of our target population 2,960 (30%) people responded. In general participants in both hospitals were satisfied with their work (mean 7.3-7.8 out of 10 maximum points), but satisfaction with the registration process scored lower (mean 5.1-6.5). The groups least satisfied with data registration were physicians, nurses, and scientific researchers (scoring < 5.5). There was no significant difference between the two hospitals in both work and registration satisfaction.

The most time spent on data registration was reported by nurses (40.2%), administrative staff (38.3%), and physicians (37.1%) the lowest was the group of facilitating staff (16.8%). We found that researchers, and analytical, facilitating, administration and support staff were willing to spend more time on registration for the sake of data reuse. Physicians and nurses wanted to spend 6.1% less time on registration with the new EHR, even when it is beneficial for reuse. In general 8.8% of the personnel (17% among specialists) expected the registration time to increase with the new EHR. About 38% do not have a clear view on future developments regarding data registration and reuse. In particular, they had no clear view on possible benefits of the EHR.

Conclusion

We found no statistically significant differences between health personnel currently working with a legacy EHR versus a patient record on paper. Nurses and medical specialists indicated they want to spend less time on data registration, even when it might save them time reusing the patient data. The majority of the participants was not satisfied with the current data registration process. Over one third of the participants did not have a clear view of what the new EHR will bring regarding data registration and reuse.

Address for correspondence

Erik Joukes: e.joukes@amc.uva.nl

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-998

Quality of Life Measurements in Spinal Cord Injury Patients

Enea Parimbelli[a], Gabriella Fizzotti[b], Caterina Pistarini[b], Carla Rognoni[c], Silvana Quaglini[a]

[a] *Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy*
[b] *IRCCS Salvatore Maugeri Foundation-FSM, Pavia, Italy*
[c] *CERGAS, Bocconi University, Milan, Italy*

Abstract

We recently developed UceWeb, an application for direct elicitation of utility coefficients (UCs), i.e. a measure of health states quality perceived by patients. UceWeb was used to interview a sample of patients affected by spinal cord injury (SCI). A standard questionnaire for measuring quality of life (QoL) and another one for the system evaluation were also administered to the same patients. The aims of this work are to (i) evaluate UceWeb usability; (ii) investigate relationships among QoL values elicited with different methods, (iii) create a reference set of UCs for the health states experienced by SCI patients. We show preliminary results obtained with the first 20 patients. Despite great variability found among QoL values elicited with the different methods, interesting correlations with patients' condition and profile have been found.

Keywords:

Quality of life, Utility Theory, Spinal Cord Injuries

Introduction

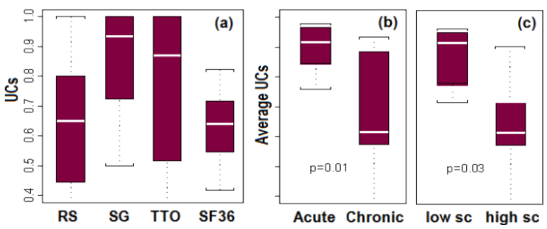
Measuring Quality of Live (QoL) of patients and correlating it with their clinical conditions may help to personalize treatments. Besides the individual level, specific QoL measures, namely the "utility coefficients" (UCs), are essential in cost-utility analysis, a technique used to evaluate interventions at a population level. UceWeb can be used by any physician, during face-to-face visits, to elicit UCs with the *classical* direct methods, i.e. rating scale (RS), time-trade-off (TTO), and standard gamble (SG) [1]. For each elicitation, a patient's (anonymous) profile is stored, so that, as the repository grows, it will be possible to retrive UCs for target populations. Moreover, relationships among UCs elicited with different methods could be studied, which also is a research issue. In this poster, we illustrate our first application of UceWeb, related to patients with SCI; an insult to spinal cord usually causing permanent and often devastating neurologic deficits and disability. With SCI, the ultimate goal of rehabilitation is to enhance QoL.

Materials and Methods

Twenty patients were consecutively recruited at the FSM Spinal Unit from October to December 2014. They were classified by the American Spinal Injury Association (ASIA) scale [2], from A (greatest impairment) to E (normal state). Using UceWeb, patients were firstly asked to rate their health state on a 0-100 scale (RS). Then they were asked to consider hypothetical scenarios: whether or not, and at which extent, they would give up part of their lifetime but living healthy (TTO); and which risk they would take, e.g. for a surgery, in order to be cured (SG). UCs were then calculated through known formulas. In addition, they answered the SF36 questionnaire, from which UCs were calculated using Brazier formula. A further Likert scaled questionnaire for evaluating the interaction with UceWeb was administered.

Results

Figure 1 – Some statistics on the elicited utility coefficients.



Boxplots show median and interquarile range (sc=scalarity)

All but 2 patients were very collaborative. Patients seem to understand RS and TTO methods slightly better than SG, for which two patients were not able to answer. While showing high variability, all scales are significantly correlated. According to past literature, RS values are significanly lower ($p<0.01$) than TTO and SG, both of which incorporate the concept of risk in their definition (Fig.1a). RS and SF36 values are similar. Chronic phase implies lower UCs than acute (Fig.1b), probably due to burden of reintegration in daily life. Low scalarity patients had higher UCs (Fig.1c). Moreover, even without statistical significance (probably due to small sample size), ASIA C patients have higher UCs than A and B, and paraplegic patients have higher values than tetraplegic.

Conclusion

UceWeb is a usable application for collecting UCs in a homogeneous way. Results obtained in SCI can be the basis for further reasoning on the effect of SCI on the quality of life.

References

[1] Parimbelli E, Sacchi L, Rubrichi S, Mazzanti A, Quaglini S. UceWeb:a web-based tool for Collecting and sharing quality of life data. Methods Inf Med. 2014 Nov 3; 53(6).

[2] Kirshblum SC, Waring W, Biering-Sorensen F, Burns SP, Johansen M, Schmidt-Read M et al. Revision of the International Standards for Neurological Classification of Spinal Cord Injury. J Spinal Cord Med. 2011 Nov; 34(6):547-54.

Address for correspondence

Enea Parimbelli, enea.parimbelli@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-999

A Methodology for Adapting Psychoeducational Content to Mobile Platforms

Stephanie Tucker[a], Sriram Iyengar[a], Amy Franklin[a]

[a] *School of Biomedical Informatics, University of Texas Health Science Center at Houston*

Abstract

Studies show that current modes of psychoeducation (PE) cannot be availed of by a substantial population of those in need. Mobile health technologies have great potential to serve such populations. However converting PE to mobile platforms is challenging. We present a methodology for this purpose based on existing learning styles theory, and developed PE apps successfully, using it. Conclusion: Useful PE apps can be developed easily using the proposed method.

Keywords:

Psychoeducation; apps; smart mobiles; rich media.

Introduction

Mental disorders affect 26.2% of American adults [1]. Although Cognitive Behavior Therapy (CBT) and Psychoeducation (PE) are effective, these are under-utilized because social (stigma), financial, and time constraints prevent many from accessing these therapies [2].

Mobile health (mHealth) technologies on smart mobiles (phones/tablets) offer great potential for overcoming these barriers. PE content on widely available smart mobiles can decrease the need for specially trained therapists. Cost of PE apps is typically less than one-on-one CBT and PE courses. Interactions with apps can be done privately, alleviating social anxieties in face-to-face settings.

However, effective adaptation of psycheoeducational content to smart mobiles is a challenging informatics problem. Existing PE content cannot simply be copied into mobiles. Greatest benefits can be achieved only by transducing content in ways that enhance usability, learning, and take full advantage of the powerful capabilities of smart mobiles.

We present a methodology, based on the Felder-Silverman Learning Style Theory [3], and prior experience with clincal practice guidelines on mobile platforms [4] to adapt PE content to smart mobiles. The methodology is generalizable to various kinds of PE content.

Methods

Adaptation of Felder-Silverman Learning Styles theory to mobiles yielded requirements that the PE content should be Active (quizzes, data intake, communicative), Reflective (self-paced), Sensing (interactive touch screens), Intuitive, Step-by-step, and have rich media (text, audio, images, video). To achieve these requirements, we converted PE content into an information flow model. We enabled sensing through touchscreens; interactivity by providing input modes such as sliders, application initiated photo and video taking. Rich media were built into every screen of the app, appealing to verbal and visual styles of learning. The typical large volume of PE content was broken up into smaller lessons in sequential order based on major content categories. In effect, the PE content was converted into a media rich patient-oriented Clinical Practice Guideline (CPG). Such guidelines on smart mobiles have high user acceptability and effectiveness in enabling adherence to guidelines [4]. Informatics challenges encountered include managing complexity of PE content, ensuring usability and navigational simplicity, choices of colors, fonts, text sizes, etc. App development was done, without programming, using the guideVue® system.

Results

Several PE apps were developed using the above methodology such as “road map to peace of mind” to provide PE to homeless veterans with mental disorders.

Conclusions

A broad, well-defined methodology is proposed to transduce face-to-face and paper-based PE to smart mobiles. It generalizes to any PE content that can follow a CPG format.

Disclaimer: The second author is part owner of a company aiming to comercialize the guideVue® system.

References

[1] Kessler, R.C., Chiu, W.T., et al. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the national comorbidity survey replication (NCS-R). Archives of General Psychiatry, 62(6), 617-27.

[2] Pietrzak, R.H., Johnson, D.C., et al. Perceived stigma and barriers to mental health care utilization among OEF-OIF veterans. Psychiatric Services, 60(8).

[3] Felder, R.M, Silverman, L.K.. Learning and teaching styles in engineering education..Engineering Education, 78(7)

[4] Florez-Arango, J.F., Iyengar, M.S.,et al.. Performance factors of mobile rich media job aids for community health workers. Journal of the American Medical Informatics Association, 18: 131-137.

Address for correspondence

Stephanie.R.Tucker@uth.tmc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1000

Special People in Routine Health Information Systems Implementation in South Africa

Lyn A Hanmer[a], Edward Nicol[a]

[a] *Burden of Disease Research Unit, South African Medical Research Council, South Africa*

Abstract

An analysis of roles and decision making structures to facilitate routine health information system (RHIS) implementation and use in public health facilities in South Africa identified a wide range of stakeholders in these processes. Two broad categories of RHIS 'special people' are analysed, i.e. leaders (administrative and/or clinical) and bridgers/support staff. In addition to health system personnel with specific responsibility for RHIS, users with an interest in effective use of RHIS and RHIS outputs, and staff of external system and/or service providers, can play significant roles in RHIS implementation and use.

Keywords:

Routine health information systems (RHIS); Health information systems (HIS); Health information system personnel; stakeholders; public sector; South Africa

Introduction

Ash *et al.* [1] have described in detail the roles of 'special people' who are essential to the success of computerised physician order entry (CPOE) systems. The focus of this analysis is the special people who contribute to the success of computerised routine health information systems (RHIS) in South Africa. The environments being considered include a wide range of public sector facilities, from primary health care (PHC) clinics to tertiary hospitals. In practice, the clinical and administrative leadership in many facilities, especially at PHC level, is provided by nurses.

Methods

The Ash *et al.* framework published in 2003 [1] identified three categories of special people: administrative leadership, clinical leadership, and bridgers/support staff. This framework has been adapted to reflect the situation for public sector RHIS in South Africa. The knowledge gained from detailed mixed-method studies of RHIS use in three of the nine South African provinces [2,3] was used to map the categories of special people identified in South African RHIS against the categories identified by Ash *et al.*.

Results

The revised framework for special people for RHIS in South Africa reflects an overlap between the personnel responsible in practice for administrative and clinical leadership, especially in primary health care settings, where the facility managers, who are responsible for administrative and clinical leadership, are likely to be nurses. Due to limited resources, including highly skilled resources to support the development, acquisition, implementation and use of RHIS, some of the roles in the Ash *et al.* framework are combined or do not exist in the RHIS environments studied.

There are multiple groups of people under the category 'bridgers/support staff'. Examples of identified RHIS special people indicate that there are additional categories of significance, including support people from outside the public health services (e.g. employed by health information system (HIS) suppliers, or by external service provider organisations to provide support to RHIS implementers). A different but important category of special people are those users with an interest in effective use of RHIS and RHIS outputs, who develop influential roles beyond those formally identified in their job descriptions, related to support for other users and/or in ensuring effective use of the RHIS in their environments.

Conclusions

These results highlight the multiple essential roles of people in management and at service delivery levels in contributing to successful RHIS implementations in South Africa.

Acknowledgments

The contributions of many special people to RHIS implementation in South Africa are gratefully acknowledged.

This study was supported by the South African Medical Research Council.

References

[1] Ash JS, Stavri PZ, Dyksra R, Fournier L. Implementing computerized physician order entry: the importance of special people. Int J Med Inf. 2003; 69: p235-250.

[2] Hanmer LA, Isaacs S, Roode JD. Factors associated with hospital information system success: Results of a survey in South Africa. In: Safran C, Reti S, Marin HF, editors. Studies in Health Technology and Informatics, Vol. 160: MEDINFO 2010. p347 - 351. September 2010.

[3] Nicol E, Bradshaw D, Phillips T, Dudley L. Human factors affecting the quality of routinely collected data in South Africa. In: Lehmann CU, Ammenwerth E, Nøhr C, editors. Studies in Health Technology and Informatics, Vol. 192: MEDINFO 2013. p788 - 792. August 2013.

Address for correspondence

LA Hanmer, South African Medical Research Council, PO Box 19070, Tygerberg, South Africa, 7505. lyn.hanmer@mrc.ac.za.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1001

Emergency Department Information System Education and Training for Clinicians: Lessons Learned

Marcelo Lopetegui[a,b], Bernd Oberpaur[a], Macarena Vivent[a], Cecilia Carrasco[a], Alejandro Mauro[a]

[a] *Clínica Alemana de Santiago, Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile.*
[b] *Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA*

Objective

Of all the potential barriers to a successful Electronic Health Record (EHR) adoption, the importance of training is often underestimated, potentially jeopardizing the implementation[1]. Following best practices recommendations, we designed and implemented a comprehensive EHR training framework. The aim of this poster is to describe our experience with such a framework in the implementation of our home-grown Emergency Department Information System (EDIS), report lessons learned and provide recommendations for other institutions facing EHR adoptions in Chile and Latin America.

Keywords:

Training; Emergency Department Information System; Education; Organizational Implementation.

Methods

Clínica Alemana de Santiago (CAS) is a 424 bed private hospital. The Emergency department (ED) receives 228,781 consults per year and has utilized a commercial EDIS since 2008. We designed and conducted the comprehensive training strategy in preparation for the deployment of a completely new home-grown EHR. We desigend a three-stage framework based on **education** (3 hours online course with teoretical material, short demo videos, and a test), **training** (45 min. 1-1 sessions where the clinician follows user stories, recording milestones completion and receiving directions if required) and **on-site support** (continuous 24/7 crew of technical and clinical experts serving as a personalised support and bug reporting system, for three months since the rollout).

Fourty five days prior to the rollout, we invited all the ED clinical staff to access the on-line education system using their institutional credentials, and once completed and approved, to schedule a training session either via e-mail, phone or walk-in (if available). Everybody was expected to complete the training before the rollout. After the rollout, reduced capacity training was still available for new hires.

Results

The described methodology provided the required preparation, contributing towards the success in the adoption of the new EDIS. We received positive feedback from unit administrators who perceived well prepared clinical staff for the implementation. Although the model was succesful, interesting lessons were learned, which will be addressed before the rollout of our new outpatient module of the EHR.

Discussion

Based on an extensive literature review and our experience implementing the framework, we have gained valuable experience including:

- Manage timing with the developer team: gain access to stable version in advance to prepare training material.
- Consider also enrolling residents in the training framework.
- Reduce size of theoretical written material and provide more short categorized videos.
- Online education needs to be differentiated: distinct courses with different coverages for distinct clinical entities.
- Demonstrate the value of your training to clinicians by showing how a well-organized chart helps to provide a safer and more efficient patient care, and the explain the relevance of good quality clinical data for secondary use.
- Deploy satellite training facilities in the users' regular resting spaces (i.e. physician's lounge).
- Provide enhanced training to key clinical stakeholders given that peer-to-peer training is usually preferred by clinicians.
- For newly hired clinicians, integrate the EHR training and education within the regular clinical service orientation classes, introducing the learner to the complete culture, processes and documentation methods of the organization.
- Avoid implementations during the summer: less crowded ED did not compensate for the amount of floating staff and the double training effort required when primary staff returned to work. Find a low demand period with primary staff available.
- Only provide access to EHR if education module is approved. However, be prepared to provide "emergency on-site training", for untrained healthcare providers showing up for their shifts withouht propper education and training
- Provide a distinctive uniform that differentiates your support staff from usual clinical staff, facilitating their location by clinicians, while preventing confused patients seeking help from them.
- Maintain a continuous streamlined feedback and bug reporting mechanism.

References

[1] Weir C, Lincoln M, Roscoe D, Turner C, Moreshead G. Dimensions associated with successful implementation of a hospital based integrated order entry system. Proc Annu Symp Comput Appl Med Care 1994:653–7.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1002

The Role of the IT Department in Information System and Organizational Redesign

Lone Stub Petersen[a]

[a] *Danish Center for Health Informatics (DaCHI), Department of Planning, Aalborg University*

Abstract

This paper is based on a qualitative study of IT organization and clinical practice over a period of 3 years supported by additional data during the last 2 years. This paper argues that redesign is central to understanding and developing both healthcare information systems and organizations. The importance of the IT organization is also stressed throughout this paper as a pertinent partner and power when considering organizational change and learning in hospitals as it can serve both as a barrier and a catalyst of change and flexibility in the organization through information systems management. Therefore it is important to consider and secure appropriate forms for redesign and learning in cooperation with the IT department.

Keywords:

Redesign, Health informatics, Information Systems Management, Organizational implementation, Organizational change.

Introduction

In order for the IT department to accommodate the needs of the hospitals there is a need for developing an understanding of the complex relationship between the IT department and clinical practice. In this paper the complex structures and the support organization in hospitals are described. Also, with the concept of redesign exemplified through the case study, I will show how IS management is an important player in the continuous change of organizations and work practices.

Methods

The main focus of the case is on different manifestations of support-work in the North Denmark Region. The data was gathered through qualitative fieldwork in the North Denmark Region including interviews, meetings and observations, supplemented by insight into a range of documents, reports and manuals. In all, 30 interviews, 12 days of observation and participation in 10 meetings and 4 workshops were conducted.

The study method was inspired by ethnomethodology and actor-network studies. Following the actors, the interviews were semi structured and explorative in nature in order to capture the differences in practice and conceptualization of support work in the different parts of the hospital.

Results

IS-Management and Ongoing Development

A central set of tasks of the IT department is called 'systems ownership' which in this paper is termed Information Systems (IS) management. IS management involves tasks and competencies related to IS support as well as systems development in the ongoing maintenance and redesign of the information systems. In this sense, IS management is an essential part of the ongoing development of systems and work practices in the hospitals.

The Work Tasks of IS-Management

The task of IS-management involves communication with user groups, vendors and other stakeholders of systems that interface with the system in question. No real model or framework is in place for this kind of work in the timeframe of the study. The numerous steps and tasks involved in managing information systems are analyzed with a focus on the themes: managing and prioritizing requests for changes and handling challenges of systems integration and dependencies.

Managing and Prioritizing Requests for Changes

The management of information systems in the hospitals involves making sure the systems are developed further in accordance with and to enable the work practices of the hospitals. This involves getting feedback from the organization on problems, errors, breakdowns and general requests for systems functionality. The importance of continuous development of information systems was apparent in many cases of the field study.

Systems Integration and Dependencies

The information systems in the hospital are not just singular and independent applications. They are integrated and dependent on other systems and technologies in many different ways in order to share data and make different systems interact in the work practices. For the CPOE at the North Denmark region, the following are integrations and dependencies:

- PAS (Patient administration system)
- The Danish Medicines Agency (lægemiddelstyrelsen)
- ApoVision – the hospital pharmacy system
- Digital bloodsugar transfer system
- The new EHR – ClinicalSuite
- New integration to the system MedOnc where cancer treatment can be shown in the CPOE
- The new shared medical card (FMK)

Conclusion

The continued redesign of organizational practices in the hospital and the need for redesigning both systems and work practices are highlighted in the case study. Hence securing appropriate and meaningful use of information technology and systems in the hospital is of key importance in order to provide quality care for patients, as these technologies are core

aspects of the hospital work practices. The input from both general and local levels are also important for the management of information systems, and the changes and development of information systems are only brought into use through good support and educating structures. Therefore the IT department plays an important, yet easily downplayed role in the redesign of the organization through the redesign of information systems.

Acknowledgments

I would like to thank the region of North Denmark for supporting the project as well as my colleagues Pernille Bertelsen and Christian Nøhr for critical comments.

Address for correspondence

Lone Stub Petersen: lonep@plan.aau.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1004

Developing an Open-Source Bibliometric Ranking Website Using Google Scholar Citation Profiles for Researchers in the Field of Biomedical Informatics

Dean F. Sittig[a], Allison B. McCoy[b], Adam Wright[c], Jimmy Lin[d]

[a] *The University of Texas School of Biomedical Informatics at Houston, Houston, TX*
[b] *Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA*
[c] *Division of General Internal Medicine, Brigham & Women's Hospital, Boston, MA*
[d] *College of Information Studies, University of Maryland, College Park, MD*

Abstract

We developed the Biomedical Informatics Researchers ranking website (rank.informatics-review.com) to overcome many of the limitations of previous scientific productivity ranking strategies. The website is composed of four key components that work together to create an automatically updating ranking website: (1) list of biomedical informatics researchers, (2) Google Scholar scraper, (3) display page, and (4) updater. The site has been useful to other groups in evaluating researchers, such as tenure and promotions committees in interpreting the various citation statistics reported by candidates. Creation of the Biomedical Informatics Researchers ranking website highlights the vast differences in scholarly productivity among members of the biomedical informatics research community.

Keywords:

Citation ranking; h-index; Bibliometrics.

Introduction

Since Eugene Garfield first created the Science Citation Index, researchers and research administrators have used citation concepts as a means of ranking the scholarly production of researchers. Over the years, numerous citation statistics have been developed in an attempt to identify measures that are highly correlated with research productivity and scientific impact regardless of: (a) scientific field, (b) length of time of the researcher in the field, (c) the influence of a few of the researcher's highly cited articles, or (d) other arbitrarily determined factors (e.g., number of articles with a specified number of citations or number of articles in journals with an impact factor $\geq$ x). We developed the Biomedical Informatics Researchers ranking website in an attempt to overcome many of the limitations of previous scientific productivity ranking strategies.

Background

With its Google Scholar website, Google has revolutionized the citation analysis comparison process by creating an online, freely available, automatically updating, scientific information resource. With the addition of the ability for researchers to create their own "my citations" page, complete with several bibliometric calculations (e.g., total citations, h-index, i-10 index), Google provides a potential method for researchers to compare themselves to other researchers.

Methods

This open-source application is written in Node.js® and built using commonly-available open source libraries. It takes as input the list of researchers and then iteratively retrieves the listing of each person's Google Scholar citation counts, the total number of citations, the year of first citation, the i10-index, and the h-index. These values are extracted based on matching the relevant elements from each page's DOM (Document Object Model) structure.

In addition to extracting raw statistics from profile pages, the application also calculates the citations/year, i-10 index/year, and h-index/year; all computed values are written into a file in JSON format, which faciliates the display as well as downstream processing by other applications.

Results

The Biomedical Informatics Researchers ranking website is available at: rank.informatics-review.com. From 16 May 2014 – 30 November 2014, 876 unique individuals accessed the site during 1,585 sessions (mean duration 1:29).

In an attempt to predict future highly cited researchers, we calculated the correlation coefficient between the h-index and total citations (r^2=0.8) and i10-index (r^2=0.93).

Conclusion

We have developed an easily searchable, interactive, automatically updating, open source, bibliometric ranking website using Google Scholar citation profiles that includes over 600 Biomedical Informatics researchers from around the world. While there are definite limitations to both using bibliometric citation analysis to measure scientific productivity, and to automatically generate lists of articles and citations, the Biomedical Informatics Researchers ranking website has already proven to be useful for a number of important tasks.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1005

Proposal for a European Public Health Research Infrastructure for Sharing of health and Medical administrative data (PHRIMA)

Anita Burgun[a,b], Dina V. Oksen[a], Wolfgang Kuchinke[c], Hans-Ulrich Prokosch[d], Thomas Ganslandt[d], Iain Buchan[e], Tjeerd van Staa[e], James Cunningham[e], Marianne L. Gjerstorff[f], Jean-Charles Dufour[g], Jean-Francois Gibrat[h], Macha Nikolski[i], Pierre Verger[g], Anne Cambon-Thomsen[j], Cristina Masella[k], Emanuele Lettieri[k], Paolo Bertele[k], Marjut Salokannel[l], Rodolphe Thiebaut[m], Charles Persoz[a], Geneviève Chêne[a], Christian Ohmann[n]

[a] Institut Thématique Multi-organisme – Santé Publique, AVIESAN, Paris, France, [b]INSERM UMR 1138 team 22 Centre de Recherche des Cordeliers, Université Paris Descartes – Sorbonne Paris Cité, Paris, France, [c]Heinrich-Heine University, Duesseldorf, Germany, [d] Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany, [e]Health e-Research Centre / Farr Institute, University of Manchester, Manchester, United Kingdom, [f]Forsker Service, Statens Serum Institut, Copenhagen, Denmark, [g]UMR912 SESSTIM (AMU-INSERM-IRD-AMSE), Aix-Marseille University, Marseille, France, [h]CNRS, UMS 3601, Gif-sur-Yvette, France, [I]CNRS/LaBRI, Université de Bordeaux, Talence, France, [j]UMR 1027 Inserm, Université Toulouse III Paul Sabatier, Toulouse, France, [k]Politecnico di Milano, Department of Management, Economics and Industrial Engineering, Milano, Italy, [l]SaReCo/University of Helsinki, Helsinki, Finland, [m]INSERM U897-Epidemiologie-Biostatistique, Bordeaux, France, [n]European Clinical Research Infrastructures Network (ECRIN), Duesseldorf, Germany

Abstract

In Europe, health and medical administrative data is increasingly accumulating on a national level. Looking further than re-use of this data on a national level, sharing health and medical administrative data would enable large-scale analyses and European-level public health projects. There is currently no research infrastructure for this type of sharing. The PHRIMA consortium proposes to realise the Public Health Research Infrastructure for Sharing of health and Medical Administrative data (PHRIMA) which will enable and facilitate the efficient and secure sharing of healthcare data.

Keywords:

Europe, data sharing, public health, research infrastructure.

Introduction and background

In Europe, data on the health of the population differs greatly in terms of availability, size and content. In Sweden and Denmark, national health databases are administered by one single authority and contain information on the entire population on diverse aspects of population health. In France, the SNIIRAM holds information for almost the entire population on health insurance beneficiaries consuming care, including reimbursed prescribed drugs. In the United Kingdom, the CPRD contains observational data from the NHS, notably primary care records for around five million patients. In Germany, the Information System Secondary Use of Routine Health Data has been newly initiated, is maintained by the DIMDI and contains administrative claims data.

Methods

Aspects of data sensitivity, heterogeneity, legal and ethical issues of sharing, quality and semantic interoperability, as well as technological solutions to ensure security, need to be considered with new concepts and rigor as to enable sharing of health data on a European level. Interoperability challenges necessitate existing solutions to require European level efforts to harmonize and ensure data sharing, given data security and ownership. ESFRI [1] is a strategic instrument to develop interoperable platforms and research infrastructures (RIs). In Sept 2014 the authors began PHRIMA, a consortium to investigate a RI for sharing medical administrative data. This RI will specifically target national databases, extendable to patient and population cohorts.

Expected benefits

PHRIMA will strengthen technological development capacity and effectiveness through metadata definitions and semantic interoperability. It strives to harmonize European medical and administrative data, and propose technical solutions for data sharing; building on national solutions and other data shareable, reusable ESFRI projects. It supports collaborative research in record linkage and anonymization. The RI will provide secure, legal solutions for large-scale epidemiological research on observational health data.

Acknowledgements

This work has received support from INSERM/Aviesan Public Health Institute.

References

[1] ESFRI-European Strategy Forum on research Infrastructures, *Strategy Report on Research Infrastructures*. 2010.

Address for correspondence

anita.burgun@inserm.fr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1006

Towards a Tool for Malaria Supply Chain Management Improvement in Rural Ghana

Lorena Carlo[a], Suzanne Bakken[a], Lena Mamykina[a], Richmond Kodie[b], Andrew S. Kanter[a]

[a] *Department of Biomedical Informatics, Columbia University*
[b] *Millennium Villages Project, Bonsaaso, Ghana*

Abstract

The maintenance of adequate quantities of antimalarial medicines and rapid diagnostic tests (RDTs) at health facilities in rural areas of sub-Saharan Africa is a challenging task because of poor supply chain management. Antimalarial stock-outs in the communities could lead patients (that need to travel long distances to get medications) to remain untreated, develop severe malaria and die. A prototype to improve the management of health commodities in rural Ghana through the visualization of current stock levels and the forecasting of commodities is proposed.

Keywords:

Malaria disease; mHealth for logistics; User centered design; Low and Middle Income Countries; Ghana.

Introduction

Malaria is a disease transmitted by the bites of infected mosquitoes. Every 30 seconds a child dies of malaria in sub-Saharan Africa [1]. There is a need for reliable healthcare data at all levels of provision. The access to timely, complete and accurate data could improve decision-making and healthcare delivery at the point of care and improve patients' outcomes [2]. Maintenance of adequate amounts of health commodities at the point of care in remote rural areas is a major barrier to the successful management of diseases in resource-poor settings. Effective information systems in health commodity supply chains are necessary to address this challenge.

Methods

The theoretical underpinning for this research work is the Sittig and Singh 8-dimension socio-technical model, which has been used to improve health information systems in different settings and stages [3]. A User-Centered Design approach was used to design a prototype to improve the management of malaria health commodities in rural Ghana. This involved focus groups, observations, interviews, and participatory design sessions. Qualitative data were analyzed using content analysis methods. Two focus group sessions were conducted. One with health workers (n=12) to collaboratively create supply chain process workflows, and one with the managers (n=6) to validate/correct workflows developed by the health workers focus group. Challenges and opportunities in the supply chain process were investigated in the focus groups as well. Observations and interviews were performed to verify the findings from the health workers' and managers' focus groups. Informants included 14 Community Health Workers (CHWs), 6 Community Health Nurses (CHNs), 3 midwives, 2 enrolled nurses, the supply chain manager, the malaria focal person, the CHW manager, the health coordinator and the eHealth specialist. A participatory design study was performed to identify and refine requirements. Using 24 participants, 14 participatory design sessions were held. CHNs and CHWs were grouped together because they interact closely in their work and have the potential to produce a richer set of requirements in a shorter period of time. A low fidelity prototype was created and evaluated with 5 CHWs, 4 CHNs, 4 midwives, 4 enrolled nurses and 3 managers, with the objective of refining requirements and obtaining feedback on the user interface.

Results and Discussion

Three primary themes emerged from the focus groups that identify the following challenges in the malaria health commodities supply chain management process: (1) malaria health commodities shortages; (2) delays in distribution of commodities; and, (3) difficulty managing manual inventory of commodities. First, there have been shortages of malaria health commodities at MVP Ghana cluster clinics. Further, it has been reported that the CHWs have run out of malaria commodities at the communities. Second, delays in the distribution of the commodities create scenarios where the CHWs borrow medication from their peers. Similar problems occur between clinics when they are unable to obtain the commodities on time. Data quality errors were reported in the collection of the inventory of commodities. The following key activities of the malaria health commodities supply chain process were observed and documented using workflows: malaria health commodities stock levels; forecasting of malaria health commodities; order of malaria health commodities; and dispensing of malaria health commodities. Future work includes evaluation of the prototype with experts and with the end-users.

Acknowledgments

Present and Past founders: SENESCYT, Schlumberger Foundation, Bill & Melinda Gates Foundation.

References

[1] WHO. World malaria situation in 1993, part I. Weekly Epidemiological Record. 1996.

[2] Blaya J, Shin S, Yale G, et al. Electronic laboratory system reduces errors in national tuberculosis program: a cluster randomized controlled trial. The International Journal of Tuberculosis and Lung Disease. 2010;14(8):1009-1015.

[3] Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care.* 2010;19(Suppl 3):i68-i74.

Address for correspondence

Lorena Carlo, Columbia University, mlc2200@columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1007

Health Interoperability into Practice: Results of the Development of a Consent Form in a Pilot Project in a Health District in São Paulo, Brazil

Guilherme Becker Sander[a], Mauro Medeiros Borges[b], José Henrique do Prado Fay[b], Denis Costa[b], Augusto Cesar Gadelha Vieira[c], Maurício Bucciolli Guernelli[c], Moacyr Esteves Perche[c], Francisco Cantarutti[a], Izolda Machado Ribeiro[b], Cleusa Ramos Enck[b], Alexandre Forte Lombardi[b], Herberth Amaral[a], Edson Dota[a], Paulo Dornelles Picon[b]

[a] Sisqualis, Sistemas de Informação em Saúde, Porto Alegre, Brazil
[b] Hospital Alemão Oswaldo Cruz, São Paulo, Brazil
[c] Datasus, Ministério da Saúde do Brasil

Abstract

Interoperability of health information systems is a centerpiece of the "E-Health" Brazilian Ministry of Health strategy. It aims to solve at least partially the health information technology puzzle that we face today. This paper describes a health information exchange pilot project in a health district of the city of São Paulo. It discusses the results of the development of an informed consent form for health information exchange. This consent form showed excellent results, with median application time of 3 minutes and with 97.8% of patients feeling fully clarified. The patients' perception when faced with options of consent to share their data is also described.

Keywords:

Medical Informatics; Health Information Exchange; Informed Consent.

Introduction

Health information Exchange (HIE) has been proposed by health care leaders as an essential tool to improve the quality and cost-effectiveness of health care, improve management of public health data and obtaining information and assist in early detection of disease outbreaks [1]. Without previous experience well described in Brazil, a consent form was developed in a pilot project in a health district of the city of São Paulo.

Materials and Methods

A consent form including applicable laws, types of consent, patients rights, consent revocation, minor patients autonomy, proxies and legal guardians, data audit and data correction was created. A draft of a consent form was elaborated. After review by a committee, the final version was released. It was submitted to patients of Mooca health district in São Paulo, between October and November 2014. We also elaborated questions to be answered by patients after the consent process to evaluate the process and the acceptability to share clinical data in an interoperable system.

Results

Recording of consent and the study on the consent process began in October 2014 and here reported is the initial experience with 1235 patients. The average time of application of consent was 3.31 minutes and the median was 3 minutes. The minimum time was one minute and the maximum was 10 minutes.

From the total, 97.8% reported receiving adequate prior information for the signature of consent. The type of consent granted is showed in table 1.

Table 1 – Types of consent

Type of consent	Number (%)
Do not consent to share	51 (4,1%)
Consent to share at any health encounter	1101 (89,2%)
Consent to share only in emergency situations	62 (5%)
Consent to share only to epidemiological control	21 (1,7%)

The most remembered risks were wrong data entered in the system, invasion of the system and information leakage.

Conclusions

Initially we were afraid that the patients would not consent. This project showed that this is not the case. Of the total initial sample of 1235 patients, 94.1% showed confidence in the system and chose to share clinical information. Our perception is that the time invested is highly rewarding and we suggest that the consent process is maintained on a large scale. Even in a relatively short period of time, 97.8% of participants felt that the process was complete and appropriate decisions were made in participating.

References

[1] Yasnoff WA, Overhage JM, Humphreys BL, LaVenture M. A national agenda for public health informatics: summarized recommendations from the 2001 AMIA Spring Congress. J Am Med Inform Assoc 2001;8:535–545

Address for correspondence

Guilherme Becker Sander. guilherme.sander@sisqualis.com.br

Av. Borges de Medeiros 2500, Sala 1407
Porto Alegre - RS – Brasil – 90110-150

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1008

Healthcare in Disasters and the Role of RFID

Samaneh Madanian[a], David Parry[a] and Tony Norris[a,b]

[a]*School of Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[b]*University of Agder, Kristiansand, Norway*

Abstract

Disasters either natural or man-made are inevitable, and therefore disaster management has always been an important function of government. Since during a disaster healthcare is often adversely affected, a lot of effort has been made in terms of researching effective responses and ways of improving the quality of delivered care to direct casualties and the rest of the community. In this regard, information technology plays an important role to help healthcare systems achieve this goal. One of these technologies that has become popular recently is Radio-Frequency Identification (RFID). This paper explores the relationship between emergency management and disaster healthcare and examines the role of RFID. It is suggested that RFID will become an integral part of disaster healthcare and a means of improving response performance.

Keywords:
Healthcare; RFID technology; Patient Safety; Disaster Response

Introduction

No society can claim immunity to disasters, thus planned disaster management – including the four phases of mitigation, preparedness, response, recovery – is an essential requirement. If enacted properly, it can lessen the harmful effects of disasters to a great extent. During disasters, healthcare plays an important part in responding to the needs of disaster casualties. In the Haiti earthquake, for example, estimates showed that the number of immediate quake-related deaths was around 200,000 and the number of injured survivors was about 300,000.

Two crucial challenges in disasters are patient identification and medical history. During disasters, medical staff deals with casualties who are unconscious and unaccompanied, and with children without their parents, none of whom can explain their current medical situation and needs, or their previous history. Patients can also be transferred among field hospitals with little more than a short medical summary written on their casts.

Materials and Methods

RFID in Today's Health Environment

The primary goal of RFID in healthcare is the enhancement of patient safety by reducing errors in patient care, including adverse drug effects, allergies, patient–medication mismatches and medication dosage errors.

RFID can reduce the inherent human errors and consequent risk by automating the communication process and delivering relevant and accurate information to responsible organizations that can employ medical resources rapidly and efficiently.

RFID in Disaster Response

The considerations noted above apply to normal, non-disaster situations, but they are even more pertinent to catastrophes where reducing loss of life and rapid recovery in the immediate aftermath of a disaster are priorities.

These attributes mean that the use of RFID technology in disasters has recently attracted a lot of attention. Central to this interest are the three fundamental characteristics of RFID: identification (who is the patient?), location (where is the patient?), and status (what is the patient's condition?).

These characteristics are of paramount importance in emergencies, as there is a 72-hour "golden" rescue period following an earthquake during which the efficiency of emergency response procedures is key to the rescue operation. If accurate and real-time information is accessible during this period, disaster management teams can coordinate more effectively the operations led by local, regional, and national organizations, and, as a result, make faster and better joint decisions.

One of the most critical tasks during this 72-hour window is triage, which not only needs to be correct and rapid, but must be recorded and communicated accurately to the responding clinicians. An RFID tag, when attached securely to a disaster victim, readily supports these requirements. The information on the tag can also avoid delays caused by repeated requests for the same information as well as be transmitted to a hospital in advance of the patient's arrival to expedite treatment.

In addition, RFID tags can monitor evacuees and control the procurement and targeting of medical supplies in order to help save lives in major disasters like earthquakes. RFID can also integrate information flow automatically, promoting its visibility among disaster responders and coordinators in different organizations, and helping them to coordinate their responses.

The benefits noted above will only be realized if several current limitations of RFID are addressed. The constraints include implementation costs, the interoperability between competing RFID standards, scaling problems, and the dual hurdles of privacy and security. These concerns are as important as the technical issues, and future research must address them in combination to earn the confidence and trust of both practitioners and patients.

Address for correspondence

Samaneh Madanian School of Computer and mathematical sciences, AUT, Auckland , New Zealand <samaneh_madanian@yahoo.com>

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1009

Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange

Brian E. Dixon[a,b,c], P. Joseph Gibson[a,d], Karen Frederickson Comer[e], Marc Rosenman[b,f]

[a] *Indiana University Fairbanks School of Public Health, Department of Epidemiology, Indianapolis, IN, USA*
[b] *Regenstrief Institute, Center for Biomedical Informatics, Indianapolis, IN, USA*
[c] *Veterans Health Administration, Health Services Research and Development Service, Indianapolis, IN, USA*
[d] *Marion County Public Health Department, Indianapolis, IN, USA*
[e] *Indiana University-Purdue University Indianapolis, Polis Center, Indianapolis, IN, USA*
[f] *Indiana University School of Medicine, Department of Pediatrics, Indianapolis, IN, USA*

Abstract

Assessment is a core function of public health. Comprehensive clinical data may enhance community health assessment by providing up-to-date, representative data for use in public health programs and policies, especially when combined with community-level data relevant to social determinants. In this study we examine routinely collected and geospatially-enhanced EHR data to assess population health at various levels of geographic granularity available from a regional health information exchange. We present preliminary findings and discuss important biases in EHR data. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

Keywords:

Electronic Health Records; Health Information Exchange; Geographic Information Systems; Community Health Planning; Health Services Needs and Demand

Introduction

Public health authorities monitor population health to identify burden of disease, manage health assets, establish policy, and evaluate interventions. This assessment usually relies on a limited set of information available through surveys, vital records, and paper-based disease reporting. Electronic health record (EHR) systems may provide more timely data for a larger portion of a population. Yet there exist a number of challenges to routine use of EHR data, including linking them to community data about social determinants of health. In this study, we sought to develop and evaluate neighborhood-level indicators of population health using EHR data integrated with a community information system (CIS).

Methods

The Indiana Network for Patient Care (INPC), a large health information exchange with over 5 billion clinical observations from EHR systems, was geospatially enhanced and combined with social determinant data from SAVI, a community information system serving the same geographical region (1). We then assessed the prevalence of diseases of public health interest and calculated HEDIS-like clinical quality indicators. Using statistical methods we assessed the reliability and representativeness of these data to measure population health at various levels of geographic granularity

Results

Rates of diabetes ranged from 1.5% to 16.07% with an average of 8.9% among neighborhoods spread across a metropolitan area. When examined at the census tract level, diabetes rates ranged from 1.5% to 12.83% with an average of 8.9% of the population for a given area.

We identified three biases in using EHR data. First, EHR data only represent those that seek health care. Second, linked EHR data are biased based how patient records were matched. The HIE uses a probabilistic technique, which can result in duplicate records. Third, the HIE proportionally contains more data from low income providers. We are exploring ways to adjust rates and correct these biases so they do not overestimate burden of disease and poor care quality in inner-city neighborhoods. We seek to compare census tract, neighborhood, and other geographic area measures with data from a recent population survey.

Conclusion

EHR systems capture data about more people than do population surveys, but they have biases that affect their estimates of population health indicators such as disease prevalence or preventative screening rates. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

Acknowledgements

This project was supported by the Robert Wood Johnson Foundation (No. 71271) by its National Coordinating Center for Public Health Services and Systems Research.

References

[1] Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. Public Health Rep. 2011;126 Suppl 3:54-61.

Address for correspondence

Brian E. Dixon, MPA, PhD, 410 W. 10th St, Suite 2000, Indianapolis, IN, 46202, USA bedixon@regenstrief.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1010

The Collaborative Coordination of Special Interest Groups on the Telemedicine University Network (RUTE) in Brazil

Thiago Delevidove de Lima Verde Brito[a], Roberto Silva Baptista[a], Paulo Roberto de Lima Lopes[b], Ana Estela Haddad[c], Luiz Ary Messina[b], Ivan Torres Pisa[a]

[a] *Graduate Program in Management and Health Informatics, Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil*
[b] *RUTE National Coordination, Rede Nacional de Ensino e Pesquisa (RNP), Brazil*
[c] *Odontology Faculty, Universidade de São Paulo (USP), São Paulo, Brazil*

Abstract

In Brazil the Telemedicine University Network (Rede Universitária de Telemedicina RUTE) is an initiative that among others promotes collaboration between university hospitals, universities, and health professionals through information technology infrastructure and special interest groups (SIGs) support. This paper presents results of analyses on collaboration during implementation and coordination activities of RUTE SIGs. This study is based on descriptive statistics and data visualization previously collected by RUTE national coordination relative to the status in July 2014. The analysis through collaboration graph identified the strongest collaboration RUTE units. The graph also highlights the collaborative relationship of RUTE units in form of communities, the most collaborative with each other in a communion in the same SIGs, and the less collaborative units in the network. It should be stated that the most active units are also the oldest in the community.

Keywords:

Telemedicine; Telehealth; eHealth; Hospitals; Teaching.

Introduction

One of the Telemedicine University Network (RUTE) [1] integration actions is to promote special interest groups (SIGs). In these groups, created and coordinated by RUTE members, institutions distributed in all Brazilian states and covering all university hospitals belonging to federal universities, health professionals plan an agenda of videoconferences and webconferences to discuss specific themes in different health specialties and subspecialties. This paper, presents results of collaboration analysis by implementing and coordinating SIG activities among the practices of telemedicine and telehealth on RUTE.

Methods

This study was developed based on previously collected data by RUTE national coordination, that also supported the production of the second RUTE book about its 100 first units launched and fully operational [1] in Brazil. During the period of data collection which lasted from April to July 2014, coordinators of the first 100 RUTE units and 60 SIGs were invited to answer an online survey about their current status. Based on these data, a process of data visualization in form of graphs to represent the social network analysis [2] on RUTE focusing on coordinating SIG units was carried out. Three main highlights were noted on the following graphs and discussed as relevant results.

Results

As shown in Figure 1, RUTE units (nodes) considered most collaborative are located in the center of the graph and are interconnected according to the SIGs coordination number. Those units are: UNIFESP, UERJ, UFBA, ISCMPA, UFSC, UFES, UFRN and HSL. Acronym: http://bit.ly/1Ozb5kK.

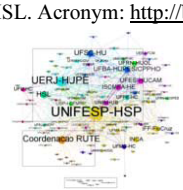


Figure 1– RUTE network graph based on SIG collaboration

The graph represents the collaborative relation of RUTE units in form of communities, where groups of nodes that have same color mean higher degree of relation and common participation in the same SIGs. Thus, we can note for example a community formed by National Coordination RUTE unit (Coordenacao RUTE). The graph represents also RUTE units with smaller collaboration degree: FioCruz-Canal Saúde, Amparo Maternal, UFPR, UFAC, HMOB, CHMSA, FMTAM, UFRJ-IG and FHAJ.

Conclusion

The analysis through collaboration graph identified the strongest collaboration RUTE units, the collaborative relation of RUTE units in form of communities, and units with lower collaboration degree.

References

[1] Messina LA, Ribeiro JL, Lopes PRL. RUTE 100: As 100 primeiras unidades de Telemedicina no Brasil e o impacto da Rede Universitária de Telemedicina (RUTE). 1ª Ed. Rio de Janeiro. 2014; pp. 504.

[2] Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge: Cambridge University Press, 1999: pp. 857.

Correspondence

Department of Health Informatics, UNIFESP, Botucatu Street, nº 862, zip code: 04023-062, São Paulo-SP, Brazil.

Thiago Lima Verde Brito, +55 (11) 5576-4522, thiago@rute.rnp.br.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1011

A Novel Approach to Teach Medical Informatics

Jürgen Holm[*], Thomas Bürkle[*], Rolf Gasenzer, François von Kaenel, Stephan Nüssli, Serge Bignens, Sang Il Kim, Michael Lehmann

Institute for Medical Informatics, Bern University of Applied Sciences – Engineering and Information Technology, Bern, Switzerland [*]contributed equally

Abstract

The Bern University of Applied Sciences established the first Swiss Bachelor of Science in Medical Informatics. We demonstrate the specifics of the new curriculum that was based on requirements analysis and includes seminars and live case studies to enable problem-solving abilities.

Keywords:
Education; Medical Informatics.

Introduction

While other countries established academic medical informatics curricula for many years [1], Switzerland did not have such a program until recently. In 2009, a survey of requirements was conducted among 22 Swiss healthcare specialists (selected by quota from hospitals, healthcare IT companies, government, suppliers of medical equipment, service providers, consulting and health insurance). The survey provided the requirements needed for 700 medical informatics graduates during the next five years in Switzerland [2].

Methods

The survey results were used to set up a detailed score for the most and least needed study subjects in a five and ten year perspective. Interoperability, clinical pathways, and the Swiss national eHealth strategy ranked as the most popular subjects for the five year perspective whereas telemedicine and a national electronic health record dominated for the ten year lookout. Based on these results, a bachelor curriculum of 180 ECTS in 49 modules (three years full time or 4.5 years part time) was implemented in 2011 and comprised of explicitly the most desired study subjects in Biological Sciences modules such as clinical pathways or interoperability.

Results

The final curriculum splits in three core components: informatics (54 ECTS); healthcare and medicine (54 ECTS); and management and communications (58 ECTS) plus 14 ECTS for optional subjects. Module topics were carefully chosen considering the survey results. Informatics is covered with courses such as basic informatics, medical statistics, methods of programming, databases, software engineering and networking. Healthcare includes courses such as anatomy and physiology, medical documentation and pathology, and also includes interoperability, accounting in healthcare, clinical studies, clinical pathways, eHealth, and quality management in medicine.

Most interesting, however, the management and communications area has increased the number of courses, and together with the healthcare and informatics area, the two groups created a series of seminars, tutorials and exercises. In these seminars, students are guided towards the application of knowledge in practical exercises and self governed work. Students receive one or two publications around a medical informatics "hot topic" like big data in medicine or personalised medication. They perform their own research including organized visits to events such as the Swiss medical fair IFAS or the German conhIT, where they can talk to international experts. They are encouraged to search interview partners in the Swiss Medical and Informatics scene and must present their results in an oral presentation, a poster, a short paper and a detailed report. Thus the other experts profit from the student work. Live cases go beyond and result in clickable mockups or functional prototypes for a medical informatics application area (often together with industrial or hospital partners). Examples of live cases include an application to identify patients for radiation therapy using barcode scanners to enhance patient safety, or the design of a hospital portal for patient admission to radiology.

Conclusion

The first graduates finished in 2014 and received immediate employment. Although many recommendations on medical informatics education exist [3], they do not consider country specific requirements, e.g., local ehealth strategies or electronic patient dossier programs. Technology and healthcare changes mandate adaptation of educational curricula. Survey-based adjustment and teaching improved problem-solving abilities seem to be promising strategies in these circumstances.

References

[1] Dickhaus HA. Comparative Summary of Six Health / Medical Informatics Programs. Methods Inf Med 1994: 33(3): 254-7.

[2] Holm J, et al. Resultate zur Umfrage über die Ausbildung von Medizininformatikern in der Schweiz. Swiss Medical Informatics 2017:70: 3-12.

[3] Mantas J, et al. Recommendations of IMIA on Education in Biomedical and Health Informatics, Methods Inf Med. 2010: 49: 105-20.

Address for correspondence

Professor Dr. Jürgen Holm juergen.holm@bfh.ch

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1012

Computer Experience of Nurses

Luciana Schleder Gonçalves[a], Talita Cândida Castro[b], Soraya Fialek[c]

[a] Nurse, Master in Health Informatics, PhD in Nursing, Professor at Federal University of Parana, Member of Research Group on Politics. Management and Health Practices GPPGPS-UFPR; [b] Nurse, Nursing Graduate Student at Federal University of Parana, Member of Research Group on Politics, Management and Health Practices GPPGPS-UFPR; [c] Nursing Undergraduate Student at Federal University of Parana, Member of Research Group on Politics, Management and Health Practices GPPGPS-UFPR

Abstract

This study aimed to identify the computing experience of nurses in southern Brazil, through exploratory survey research. The results, which were obtained from the application of The Staggers Nursing Computer Experience Questionnaire®, were analyzed by statistical tests. The survey was conducted with nurses working both in hospitals, as in public health, in a capital in southern Brazil. There is the predominance of novice nurses in the application of computer tools in their practices but most often declare the use the computers to develop their professional and also personal life activities. We conclude that the computer and health information systems are part of the working reality of the participants, being considered indispensable resources his activity, while noting limitations on the potential use of these tools. This study reflects on how the issue has been addressed in educational schools and the challenges of inclusion of the theme of Nursing Informatics in the curricula in Brazil.

Keywords:

Health Information Systems; Nursing Informatics; Nursing Informatics Competencies.

Introduction

The incorporation of new technologies has occurred steadily over the evolution of nursing, a fact that has enabled the expansion of the professionals' practice. It could have contributed to the improvement of the scope of their professional practice. Nurse's performance in different scenarios: caring, education, management, and research. Information and communication technologies (ICT) are support tools that can help health professionals achieve the efficient management of the high range of data produced in healthcare area (1). In 2013, the Center for Studies on Information and Communication Technologies (CETIC.br) promoted a large and pioneering study on the use of ICT in health facilities in Brazil, as well as appropriation of these technologies by industry professionals - the "Health ICT" research (2). In general, this research shows that professionals see as positive the implementation and use of information in health systems. They point out that the inclusion and use of these tools add value to the work process, thus contributing to gains in terms of efficiency and quality. It is believed, however, that the professionals are often placed in the context of ICT in the workplace, but without having received the necessary training to their best use. This can cause them to have little positive view of the tools as a means of facilitating their work (2).Based on these assumptions, this study aimed to identify the computing experience of nurses among the nursing community of Curitiba, Paraná.

Methods and Results

This is a survey-type survey, exploratory, with a quantitative approach that uses the descriptive statistics for the analysis of its results. The project was approved by the Research Ethics Committee. 78 nurses filled out the questionnaire at the University Hospital and 260 filled at the Public Health Network. Participants are endowed with both experiences as specific knowledge in any field of nursing knowledge (most have specialist titles, and some are masters and doctors). Overall, they reported the use of computers more for administrative purposes, and less to clinical issues and nursing care records. Participants' responses reveal that their majority does not have formal knowledge in computer science, but declare to have done quick courses on computer applications and read magazines or books on computing.. However, most rated their computing experience to be competent. Most of them declare to use computers often.

Conclusion

Knowing the computing experience of nurses is the first step to an awareness of the presence of ICT in health care. It is a reality that can no longer go back, but also an achievement of individual and organizational professional development goals in health.

References

[1] Gonçalves LS. Competências em informática requeridas de enfermeiros na prática profissional brasileira [tese]. Universidade Federal do Paraná, Curitiba, 2013.

[2] Brasil. Comitê Gestor da Internet no Brasil (CCGI.br). Núcleo de Informação e Coordenação do Ponto BR (NIC.br). Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC.br). TIC Saúde 2013 – Pesquisa sobre o uso das TIC em estabelecimentos de saúde no Brasil. Coletiva de Imprensa – Apresentação de Resultados. São Paulo, dez. 2013. Available: <http://www.nic.br/imprensa/ releases/2013/rl-2013-65-ppt.pdf> Acesso em: 10 fev. 2014.

Address for correspondence

Luciana Schleder Gonçalves
Av. Lothario Meissner, 632 Bloco Didático II
Curitiba – PR – Brazil 80210-170
lualevale@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1013

Informatics Competencies in Nursing Management

Rodrigo Jensen[a], Christiane Pereira Martins Casteli[bc], Rika Miyahara Kobayashi[c], Maria Madalena Januário Leite[b]

[a] *Botucatu Medical School, Sao Paulo State University (UNESP), Botucatu, Brazil*
[b] *School of Nursing, University of Sao Paulo, Sao Paulo, Brazil*
[c] *Dante Pazzanese Cardiology Institute, Sao Paulo, Brazil*

Abstract

This study aimed to identify in scientific literature the informatics competencies required from the nurses to make decision in management process. Through a scoping review, literature databases were searched to find articles published in Portuguese, English, or Spanish, until July 2013. 188 articles were found, and seven were included in this study, published between 1994 and 2011. The studies were written in English (5; 71%), in USA (5; 71%), using experience reports or literature review designs (5; 71%). The informatics competences were categorized according the Technology Informatics Guiding Education Reform (TIGER). The findings highlight gaps in informatics competence to make decisions in the management process – essentially in information management competence.

Keywords:
Nursing Informatics; Nursing Administration Research; Professional Competence.

Introduction

The use of computer technologies by nurse managers is increasing exponentially. The aim of this study was to identify in the informatics competencies required from the nurses to make decisions in the management process.

Methods

The scoping review method was used to answer the research question: What informatics competencies are required from nurses to make decisions in the management process? Articles published until July 2013, described in Portuguese, English and Spanish, in CINAHL, Embase, LILACS and PubMed databases were searched. We found 188 articles related to decision-making in the management process developed by nurses supported by informatics competencies. Articles with educational or clinical foci were excluded. In the first round, titles and abstracts were evaluated and 172 articles were excluded. In the second round, 16 full-text articles were read and evaluated by peers, and nine were excluded. Seven articles included were analyzed and the informatic competences were categorized in Basic Computer Competencies, Information Literacy and Information Management, according to the TIGER initiative[1].

Results

Seven articles published between 1994 and 2011 – predominantly in English (5; 71%), from USA (5; 71%), and with low level of evidence, using experience report/literature review designs (5; 71%) – were included. The following informatics competencies and skills were required of the nurse managers:

- Basic Computer Competencies: skills to use the Internet, spreadsheet, graphs, social networks, and databases;
- Information Literacy: skills to use minimum data sets, telemedicine, and evidence-based practice to decision-making, and ensure the quality of information;
- Information Management: skills to use algorithms, decision support systems, and information systems; and to use systems that can maximize clinical outcomes and reduce costs.

Conclusions

In the literature, discussions about how informatics competencies can support the decision-making of nurses in management process are still insufficient. The findings highlight gaps in informatics competence to make decisions in the management process, and essentially in information management competence. The articles showed the importance of infrastructure and education processes to develop informatic competences in health.

Acknowledgments

To Brazilian National Council for Scientific and Technological Development (CNPq) (n. 151306/2013-3).

References

[1] Technology Informatics Guiding Education Reform. The TIGER initiative: informatics competencies for every practicing nurse: recommendations from the TIGER collaborative 2014. Avaiable in: http://www.thetigerinitiative.org/docs/TigerReport_InformaticsCompetencies_001.pdf

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1014

Health Informatics Competences for eHealth: What Can We Learn From a Bibliometric Analysis?

Peter Kokol[a], Helena Blažun[b], Kaija Saranto[c]

[a] *Univerity of Maribor, Faculty of Electrical Engineering and Computer Sciences, Maribor, Slovenia*
[b] *Univerity of Maribor, Faculty of Health Sciences, Center for International Cooperation, Maribor, Slovenia*
[c] *University of Eastern Finland, Department of Health and Social Management, Kuopio, Finland*

Abstract

The appearance of eHealth adds a new dimension to health informatics competencies – they are not necessary just for health providers and health information system users and developers, but also for health consumers.

Keywords:

eHealth; Bibliometric mapping

Introduction

eHealth has gained a lot of attention on different fronts and hence a need for new health informatics competencies has arisen. Kokol et al. **[1]** had performed a bibliometric analysis about nursing informatics competences. The aim of this paper is to generalize this kind of analysis to other healthcare professionals in the context of eHealth.

Material and methods

Two corpuses were formed from the Scopus and Elsevier bibliographic databases. HICSC was searched using "(*medical OR health OR nursing OR clinical) AND informatics* AND (competency* or skill*)) strings.". EHCSC was searched using the search string of "(*medical OR health OR nursing OR clinical) AND informatics* AND (competencies* or skill*)) AND eHealth". Corpus data were recorded in an Excel document; preprocessed using built-in Excel functions and then transferred into VosViewer [2] for bibliometric mapping.

Results and discussion

The search was performed on 16th of November 2014. The HICSC consisted of 1779 and EHCSC of 85 information sources. The trend in both scientific areas is positive, but steadier in health informatics than in eHealth competences/skills. In health informatics competences/skills, the exponential trend started in 1990 and became steady in 2004. In the same year the exponential trend in eHealth competencies/skills started. The delayed start of research literature production in eHealth competences in regard to health informatics competences is congruent with the more intensive start of the eHealth research literature production in general.

The information sources from health informatics competences were written by authors affiliated in 86 different countries and from eHealth competences in 29 countries. The vast majority of papers were published in the United States, followed by United Kingdom, Canada and Australia. Bibliometric mapping produced following clusters from HICSC abstracts:

- **Information cluster** combining information with evidence, intervention and patient;
- **System cluster** linking system with the hospital, database and computer;
- **Nursing cluster** associating nursing with informatics competency and computers skill;
- **Education cluster** combining medical/ health informatics and electronic health record with health care and health professionals

and five clusters from EHCSC abstracts:

- **Nursing cluster** linking nursing practice with competency and electronic health record;
- **Education cluster** combining medical education with medical/ health informatics and information technology;
- **Technology cluster** associating technology with woman and smartphone;
- **Internet cluster** linking Internet with health information and college student and
- **Disease cluster** combining health care, training, diagnosis with computer science.

Comparison of HICSC and EHCSC clusters reveals a considerable difference between the research topics in the two fields of our study. The eHealth competence/skills field is oriented toward Internet, technology and diseases and the health informatics competences/skills field is directed more to systems and information.

Conclusion

The present study highlighted that one of the central concepts in health informatics research, namely electronic health records is placed in different contexts within both studied areas. This is an important discovery leading to the conclusion that to achieve the goals of patient centred paradigm, research endeavours in both studied fields should be united.

References

[1] P. Kokol, H. Blažun, J. Vošner and K. Saranto, "Nursing informatics competencies: bibliometric analysis," *Studies in Health Technology and Informatics,* vol. 201, pp. 342 - 348, 2014.

[2] N. J. Van Eck in L. Waltman, VOSviewer Manual, Leiden: Univeristeit Leiden, 2013.

Address for correspondence

Peter Kokol, Faculty of electrical engineering and computer science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia.
peter.kokol@um.si

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1015

Education, Technology and Health Literacy

Kurt Lindgren[a], Ina Koldkjær Sølling[b], Per Carøe[b], Kirsten Siggaard Mathiesen[b]

[a] *Department of Technology and Business, University College of Northern Denmark, Denmark*
[b] *Department of Nursing, University College of Northern Denmark, Denmark*

Abstract

The purpose of this study is to develop an interdisciplinary learning environment between education in technology, business, and nursing. This collaboration creates natural interest and motivation for welfare technology. The aim of establishing an interaction between these three areas of expertise is to create an understanding of skills and cultural differences in each area. Futhermore, the aim is to enable future talents to gain knowledge and skills to improve health literacy among senior citizens. Based on a holistic view of welfare technology, a Student Academy was created as a theoretically- and practically-oriented learning center. The mission of the Student Academy is to support and facilitate education in order to maintain and upgrade knowledge and skills in information technology and information management related to e-health and health literacy. The Student Academy inspires students, stakeholders, politicians, DanAge Association members, companies, and professionals to participate in training, projects, workshops, and company visits.

Keywords:

Informatics; innovation; education; training; interdisciplinary; health information technology; m-health; e-health; healthcare; health literacy; participatory health.

Introduction

Population aging leads to considerable pressure on health expenditures because elderly people are living longer with chronic diseases. There is an assumption that welfare technology can provide cost-effective solutions and have a positive impact on the participatory health of elderly people. Training and education of health care professionals is needed to ensure they possess skills to meet future health care demands [1].

Methods

The Student Academy is an interdisciplinary learning environment with a holistic understanding of technology based on the Leavitt-Ry organizational model [2].

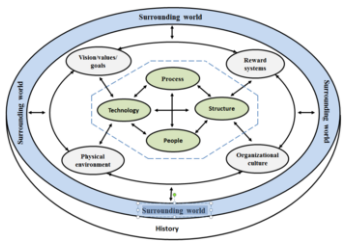


Figure 1: Leavitt-Ry model of organization [2].

The learning and training focus on a user-driven and innovative approach to welfare technology. Students from technology, business, and nursing upgrade their professional knowledge and skills in information technology and information management. Companies and health professionals participate in theoretical and practical training, projects, workshops, and company visits. A collaboration was established with DanAge Association (NGO) to develop competencies in e-health and health literacy. 114 students, 18 teachers, four companies, politicians and six from DanAge have participated in the Student Academy.

Results

Based on observations and feedback the knowledge and skills of the students have improved. The interdisciplinarity and the collaboration with health professionals' companies and the DanAge Association has estabilised an understanding of how technology can support the participatory health of elderly people with chronic disease. In an innovative process, six student groups developed a company's business plan for a new robot designed for rehabilitation.

Conclusion

An interdisciplinary learning environment was established between education in technology, business, nursing, stakeholders, politicians, DanAge Association, and companies. This collaboration created innovation, talents, knowledge, and skills related to e-health and health literacy among senior citizens.

Acknowledgments

Thanks to: KMD, DanAge Association, Future Nursing Home and Center for Welfare Technology in the Municipality of Aalborg, MEDEI Aps, E-MERGENCY, LT Automation, TeleCare North, and University College of Northern Denmark

References

[1] KL. Det nære sundhedsvæsen. 2012;18-28, 76-80. http://www.kl.dk/ImageVault/Images/id_53357/scope_0/ImageVaultHandler.aspx, Last visited: 14-12-2014, 15:30.

[2] Ry Nielsen JC, Jens C. Anderledes tanker om Leavitt – en klassiker i ny belysning. 2002. København, Nyt fra Samfundsvidenskaberne.

Address for correspondence

Kurt Lindgren, MSE
Sofiendalsvej 60
9200 Aalborg SV, Denmark
Mail: krl@ucn.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1016

Current Status for Teaching Nursing Informatics in Denmark, Canada, and Australia

Inge Madsen[a], Elizabeth Cummings [b], Elizabeth M. Borycki[c]

[a] *Centre of Clinical Guidelines, Department of Health Science and Technology, University of Aalborg, Denmark*
[b] *School of Health Sciences, University of Tasmania, Tasmania, Australia*
[c] *School of Health Information Science, University of Victoria, Victoria, BC, Canada*

Abstract

Nursing schools in Denmark, Canada, and Australia are all currently involved in integrating nursing informatics in the nursing bachelor programme. This paper gives a brief update on the current situation of nursing informatics education for bachelor level nurses in each of the three countries. Whilst there are differences in the curriculum in each county, it is important to share knowledge about undergraduate nursing informatics worldwide to ensure consistency.

Keywords:

Nursing informatics, bachelors programme, curriculum

Introduction

The explosion in information technology systems in the healthcare setting has resulted in a transformation of work practices. Nurses therefore need knowledge, skills, judgment, and understanding of the importance of informatics from the commencement of their training. This paper is a brief description of the current uptake of nursing informatics (NI) for undergraduate students in Denmark, Canada, and Australia.

Materials and Methods

A review of the published and grey literature was conducted. More specifically, the authors searched the following databases: CINAHL and Medline using the terms: "nursing", "informatics", "Canada", "Denmark" and "Australia". Grey literature was also reviewed.

Results

Nursing informatics is a topic area that needs to be integrated into the nursing curriculum internationally; however, each country has achieved different levels of integration of informatics into their degrees.

Denmark The order of nursing curriculum describes that the program in nursing informatics is to qualify the students in the theoretical and clinically technological development and to meet the people's need for nursing informatics. [1] Nursing informatics is placed in the bachelor's program, 18 months after start and holds two 2 ECTS – points (27 lessons).

Canada In 2012, the Canadian Association of Schools of Nursing (CASN) and Canada Health Infoway developed Nursing Informatics Competencies for Entry-to-practice Registered Nurses [2] and learning tools and resources that can help faculty to with teaching students undergraduate nursing informatics (see 2,3). CASN is actively involved in supporting faculty in a peer to peer network to help faculty master nursing informatics competencies and integrate them into nursing curricula across the country. Peer leaders are engaging nursing faculty across the country [3].

Australia In 2012 the Australian Nursing and Midwifery Accreditation Council released new standards for accreditation of nursing education. These standards include the requirement that student nurses have "familiarity with health informatics, including person-controlled electronic healthcare records" [4]. The importance of developing "the capacity to innovatively use information technology and electronic resources to research the growing evidence base for improved care and treatment methods" [4] is also acknowledged. As a result of this upgrade to the standards it is not possible for a nursing degree to be accredited in Australia unless it includes informatics.

Findings

Differences in the development, evolution and integration of nursing informatics into undergraduate education between the countries exist. Denmark has integrated nursing informatics successfully into the bachelor program. Canada and Australia have recently commenced this process. It is evident that the sharing of information about best practice in curriculum development in integration of informatics will enhance the uptake of this important area in undergraduate degrees.

References

[1] Danish Ministry of Education: Law about nursing education in Nursing Curriculum in Denmark. (In Danish).

https://www.retsinformation.dk/forms/R0710.aspx?id=114493

[2] L. Nagle, E. Borycki, L. Donelle, N. Frisch, K. Hannah, A.Harris, A. Harris, S. Jette, T. Shaben. Nursing Informatics: Entry-to-practice Competencies for Registered Nurses.CASN, 2012: Ottawa.

[3] CASN. Peer leaders for a nursing faculty eHealth Network. http://www.casn.ca/en/E-HealthNursing_Informatics:_CASN-Infoway_Nurses-in-Training_Project_182/items/7.html

[4] ANMAC, 2012, Australian Nursing and Midwifery Accreditation Council Registered Nurse Accreditation Standards http://www.anmac.org.au/accreditation-standards

Address for correspondence

imad@via.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1017

Trends of Patient Safety Topics Addressed in the Past Five MEDINFO Congresses

Qiyan Zhang [a], Hiroshi Takeda [a]

[a] *Graduate School of Health Care Sciences, Jikei Institute, Osaka, Japan*

Abstract

Health informatics has been promoted as a core curriculum in our patient safety program for four years. There is a need to let students understand how such topics were addressed with relevant literature of trends in medical informatics. In this study, we performed a content analysis by searching the past five MEDINFO Congress Proceedings. We used MeSH Terms or Keywords to search the study subjects and divided them into three categories. 76 subjects were identified, among which 16 were assigned to Category 1 (System-related), 26 to Category 2 (Database-related), and 34 to Category 3 (Others). Some socio-techical issues that emerged in Category 3 presented as future research interests.The identified topics reflect research trends in patient safety.

Keywords:
Patient Safety; Education; Medical Informatics.

Introduction

Health informatics has been promoted as a core curriculum in the Jikei Institute for four years.[1] There is a need to let students understand how informatics topics were addressed with relevant literature regarding trends in eHealth in general. A recently published systematic overview indicates that there is a large volume of work studying the impact of eHealth on the quality and safety of health care.[2] "What is the perspective of medical informatics within this field?" should be asked to help our students to gain domain knowledge. In this study, we performed a content analysis through checking the related topics presented in the past five MEDINFO Congress.

Methods

Pubmed's Advanced Builder was used to search the study subjects. We used Text Words – e.g., "safety" or "error" or "accident" or "incident" or "adverse" or "risk management" – to identify patient safety topics published in the "Journal of Studies in Health Technology and Informatics," volumes 84, 107, 129, 160, and 192. Then, we screened MeSH terms (Volumes 84, 107, 129, and 160) or Keywords (Volume 192), which each contained identified subjects with the aforementioned Text Words to make reconfirmation. Next, we divided the refined subjects into three categories, e.g., Category 1 (System-related), which included MeSH Terms or Keywords like EHR, HIS, Computerized Medical Record Systems, Medical Order Entry System and text words containing "system"; Categoty 2 (Database-related), which included MeSH Term or Keywords like Artificial Intelligence, Knowledge Base, Natural Language Processing (NLP), Data Mining (DM), Information Storage and Retrieval, Decision Support System or Technique; Category 3 (Others), for those topics that belong to neither Category 1 nor Category 2.

Results

In total, 76 subjects were identified, among which 16 were classified in Category 1, 26 in Category 2 and 34 in Category 3 (Table 1). 31 out of 76 subjects were medication safety related, of which 21 are in Category 1 and Category 2, and 10 are in Category 3, respectively. Some socio-technical issues (in Category 3) – for example; quality assurance, quality control, task performance analysis, practice management, workload, interdisciplinary communication, and inter-professional relations – emerged as new foci in this field.

Table 1– Patient Safety Topics Identified in the Past Five MEDINFO Congress Proceedings

Category	MEDINFO 2001	MEDINFO 2004	MEDINFO 2007	MEDINFO 2010	MEDINFO 2013
Category1	0	5	6	1	4
Category2	1	4	5	10	6
Category3	0	5	3	10	16

Discussion and Conclusion

Patient safety has been an increasing concern in the field of medical informatics during the past fifteen years. Medication safety was highly addressed on the implementation of computerised physician order entry (CPOE) systems, and application of decision support technigues to prevent error and manage risks. DM- and NLP-related methods were widely applied to detect adverse drug events. Some socio-techical issues emerged in Category 3, which presented as research interests that need to be explored further. Our findings also showed that evidence for the clinical benifical impact of such technologies is still limited. Patient outcomes in particular are seldomly being assessed in these studies. Academics in health informatics should raise more awarenesses to such problems. Future eHealth technologies are expected to be truly benifical to patients in terms of quality and safety.The topics reflect research trends and can support education in this field.

References

[1] Zhang Q, Takeda H. Promoting Health Informatics for Multidisciplinary Education in Patient Safety: A Master Course Program in Japan. MEDINFO 2013. Doi: 10.3233/978-1-61499-289-9-992.

[2] Black AD, Car J, Pagliari C et al. The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview. PLoS Medicine; January 2011; Volume 8; Issue 1; e1000387.

Address for correspondence

Dr. Qiyan Zhang (e-mail address: qyzh.louise@gmail.com)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1018

Social Network and Health Researchers and Professionals Mobility in Africa: Lessons Learned from AFRICA BUILD Project

TRAORE ST[b], ANNE A[a, b], A. KHALIFA[c], S. BOSOMPRAH[d], F. CAROLINE[e], A.K CUZIN-KIHL[f], B. INGELBEEN[g], M.RAMIREZ-ROBLES[h], SANGARE M[b], NIANG M[b], BAGAYOKO CO[a, b]

[a] *Centre d'Expertise et de Recherche en Télémédecine et E-Santé*
[b] *Faculté de Médecine, de Pharmacie et d'Odonto-Stomatologie de Bamako*
[c] *Ministry of Communications & Information Technology (ITI-MCIT) of Egypt*
[d] *University of Ghana,* [e] *University Hospital of Geneva,* [f] *World Health Organization*
[g] *Institute of Tropical Medicine (ITM) of Belgium,* [h] *Universidad Politecnica of Madrid (UPM) of Spain*

Abstract

***Objectiv**e: Promote mobility between South-South and South-North for improving level of researchers, staff and students through a platform.* ***Methods**: The methodology is based a filling of a questionnaire about offer or demand. Material is composed a computer connected Internet.* ***Result:** we registered about 203 demands and 31 offers from partners.43 mobilities were executed completely.* ***Conclusion:** The results indicate a real need of mobility for researchers and health professionals in Africa. The important number of mobility demands made by external researchers and professionals (from outside the AFRICA BUILD Consortium) may be constrained by the difficulty to find adequate funding.*

Keywords:

Personel mobility; Health; Africa Build Project.

Introduction

Scientific research is the key to change as it is on research's validated outcomes and reflections on reality that policy makers can base their decisions. At university, scientific research allows students to improve their knowledge [1].

One of AFRICA BUILD project objectives is to facilitate the mobility between learners.

Methods

The hardware consists of a platform via a computer connected to the Internet. This platform had a demand or offer form (depending on the chosen necessary) that the interested have to fill [3].

Definition of mobility

In AFRICA BUILD, mobility (or exchange programs) was defined as the exchange of students and other forms of personnel mobility aiming to conduct a training or research task in other organization.

Results

Forty three (43) mobility initiatives were completed including cases of North-South, South-North and South-South mobility (table 1). Health informatics was the more representative field of interest for this mobility with a total of 13 cases (Table 2).

***Table 1**: Completed mobilities by type*

Mobility type	Number of mobilities
North-South	12
South-North	14
South-South	9
Other*	8
TOTAL	43

***Table 2:** Completed mobilities initiatives by fields*

Fields	Frequencies	Percentages (%)
Health Informatics	13	30.23
Public Health	3	6,98
Elearning	9	20,93
Other	18	41,86
Total	43	100

Conclusion

The results of this work show that mobility is an interesting alternative for capacity building of individuals and strengthening institutions.

Acknowledgments

This study was conducted as part of Build Africa, funded by the FP7 program of the EU.

References

[1] Leclerc-Olive, M., G. S., & Wagner, A. C. (2001). *Les mondes universitaires face au marché: circulation des savoirs et pratiques des acteurs.* KARTHALA Editions.

[2] Ramirez-Robles, M., Jimenez-Castellanos, A., Khalifa, A., Anne, A., Kamga, Y.N., Afagbedzi, S., Maojo, V.: *AFRICA BUILD Portal: Developing A Social Network Of African Health Researchers And Educators.* Proceedings of IST-Africa 2013 Conference, Nairobi, Kenya. (May 2013). ISBN: 978-1-905824-38-0

Address for correspondence

Dr Seydou Tidiane TRAORE (Department of teleexpertise)
Phone : (+223) 66 71 27 83 B.P: E3791 Bamako, Mali
Email: trawore@certesmali.org www.certesmali.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1019

Middleware Supporting Next Generation Data Analytics in Australia

Douglas I. R. Boyle[a]

[a] *Health and Biomedical Informatics Centre, The University of Melbourne, Australia*

Abstract

In Australia, as a result of the distributed, often private nature of health provision, tight privacy legislation, even tighter organizational policies, access to data covering the whole patient journey of care is a common aspiration that has been almost impossible to achieve. Access to primary care data in a manner that is record-linkable has been a particular challenge. Since 2006 The University of Melbourne has been developing GRHANITE™ Middleware and GRHANITE™ Data Linkage technologies designed to overcome these barriers. With over 10% of Australian primary care data now being routinely extracted utilising this technology, we believe the principal technical challenges have now been overcome. We believe this technology to be at the forefront ethically of providing data for research. This poster describes the principal issues involved and the approaches taken in the technical solution underpinning GRHANITE™.

Keywords:
Data Integration; Primary Healthcare; Linked health data.

Introduction

Australia is a naturally conservative country with conservative privacy, security and consent policies. There are over 7,000 primary care businesses each of whom need to separately grapple with issues of the law, privacy and consent. With no central policies, accreditations or regulations, practices are conservative – and linkable data for research has been nearly impossible to obtain. The GRHANITE middleware system is a major initiative by the University of Melbourne to address the issues of data linkage and to address the many logistical issues in ethically obtaining and utilising the big hole in Australian Health Data Analytics – Primary Care.

Methods

Australian GP systems exhibit no standards in technology, almost no standards in data representation or coding and no accreditation processes. GRHANITE™ deals with these issues through its generic ability to link to almost any technology – and to extract data in raw form allowing interpretation to happen later. This helps resolve problems regarding a lack of IT knowledge in the practice. GRHANITE™ implements opt-in, opt-out and waiver of consent mechanisms - unusual for middleware and project-specific ethics requirements, aggregations and rules can be applied before data transmission. These are the principal components that help reassure practices regarding GRHANITE™ implementations. Study protocols and software can be remotely updated lowering support overheads as is necessary in implementations spanning 100's sites. A privacy-protecting hashing algorithm allows data to be linked to non-GP data sources allowing linkage to happen in this sector for the first time without opt-in informed consent. All communications are encrypted. Delta processing means only data changes are exported lowering bandwidth overheads. The poster shall explain the principal GRHANITE components and architecture.

Results

GRHANITE has been implemented in over 650 sites across Australia and interfaces to eight of the principal GP systems covering over 95% of the Australian GP market. Approximately 40 other interfaces have been developed crossing other health sectors. The collections now have the majority of the GP medical record for nearly 10% of the Autralian population. Initial studies utilizing GRHANITE are now reaching maturity.

Discussion and Conclusion

The poster will describe in detail the many challenges that need addressed to ethically extract data from primary care in Australia – a complex task with over 20 man-years development to achieve. A fast growing level utilisation of GRHANITE across the sector helps validate the mechanisms undertaken to address the challenges present – and in doing so formalises the challenges. The GRHANITE record linkage tool is probabilistic and this generally rules-out clinical data sharing applications including acting as a Health Information Exchange. Because the data exported is generic and raw, the use of GRHANITE requires good analytic skills – BI tools and data interpretation are not roles GRHANITE fulfills.

Many countries may be able to implement systems with fewer legislative and policy barriers. The fact that this tool is able to navigate so many barriers suggest to us that the GRHANITE system is a leader in the field of ethical large-scale data provision for audit, research and health surveillance. Even where legislation or policy restrictions allow more flexibility, a GRHANITE implementation may help raise the bar.

References

References will be available on the poster - Please see also **http://health.informatics.unimelb.edu.au**

Address for correspondence
Dr Douglas Boyle Ph.D., FACHI, Director, Research Information Technology Unit (HaBIC R^2), Health and Biomedical Informatics Centre, The University of Melbourne
Level 1, 202 Berkeley Street, Parkville, 3010 Victoria, Australia, Tel: +61 (3) 5823 4521
Email: dboyle@unimelb.edu.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1020

Proposal of Local Automatic Weighing Attribute in CBIR

David Jones Ferreira de Lucena[a, b], Marcelo Costa Oliviera[a, b], Aydano Pamponet Machado[b]

[a] *Lab of Telemedicine and Medical Informatics at University Hospital HUPPA/UFAL/EBSERH*
[b] *Computing Institute, Federal University of Alagoas, Brazil. {davidjones, oliveiramc, aydano}@ic.ufal.br*

Abstract

Lung cancer is the most common malignant lesion and the principal cause of cancer-related death worldwide. This problem encourages researchers to build computer-aided solutions to help diagnose lung cancer. Content-based image retrieval (CBIR) systems are very promising in this context due to a large number of image generated everyday. However, semantic gaps have limited CBIR applicability. This work proposes a new approach to automatically adjust CBIR attribute weights to reflect users' semantic interpretation on retrieval process, minimizing the semantic gap problem and improving retrieval accuracy.

Keywords:

Content-based image retrieval; information retrieval; decision support; update weighing attributes; lung cancer.

Introduction

Currently, lung cancer is the principal cause of cancer-related death. Some characteristics like the aggressive and heterogeneous nature, small size (less than 3mm in initial stages) and location in anatomic complex structures have imposed difficulty for mortality reduction by early diagnosis.

In the last decade, techniques based on computer vision, image processing and pattern classification have been applied to assist in the diagnosis process, providing a second opinion to radiologists. CBIR is one of the most promising approaches in this context due to the large amount of images generated. CBIR is a framework that could provide support retrieving images based on image content or on metadata associated to an image. Retrieving similar exams to a new one based on similarity criteria allow specialists to check another exams to assist them on decision making. Despite this, CBIR has a semantic gap problem. On the other hand, there is a gap between the users' subjectivity and feature used by computational algorithms to represent objects mathematically.

In order to minimize semantic gap, we propose an automatically updated weights of features based on local learning, aiming to represent users' sermantic interpretations and retrieve nodules with same malignancy and visually similar.

Materials and Methods

The image database utilized was the Lung Image Database Consortium (LIDC). It has two important properties of the nodules: margin manually segmented and malignancy associated to each nodule. In previous work[1], the preprocessing phase performed aimed to extract 36 texture attributes (TA) of each nodule based on a 3D co-occurrence matrix.

The proposed processing began normalizing the TA database and divided it into two balanced parts: a training and evaluation database. Next, sequential training and evaluation steps where the evaluation function (EF) increased (Figure 1). When the process stopped, we found that the weight vector learned from subjectivity in malignancy determination.

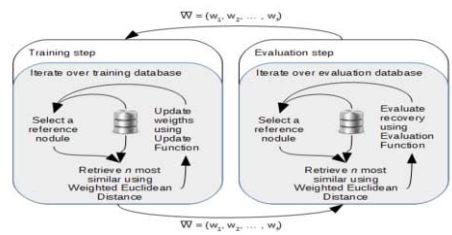


Figure 1 - Update weight process workflow

The local property of this proposal is based on using only the *n* most similar nodules to update weights, compared to utilizing all examples from a dataset for global methods. Considering its neighborhood, the weights were updated as shown in Equation (1). In order to accompliish, we calculated the inverse pattern deviation (IPD) of attribute projection a_i from *n* most similar nodules recovered by the weighted euclidean distance (WED). IPD has been used to minimize the variability in a_i, because intuitively as the variability increases, the relevance of attribute a_i decreases, and vice versa. The update rate determines how much change in each iteration.

$$w_i' = w_i + \alpha\left(\sigma_i^{-1} - w_i\right) \quad (1)$$

During the evaluation process, retrieval order and relevance of nodules was very important. The EF used an amortization rate ( ) to give more influence to top positions and assigned a score value (*s*) associated to nodule relevance, as show in Equation (2).

$$E(\overline{w}, \overline{x}) = \sum_{j=1}^{n} \gamma^j s_j \quad (2)$$

Conclusion

With this work, we aimed to contribute to the advancement of CBIR, giving an automatic solution to adjust attributes' weight in order to improve retrieval precision of similar objects and to minimize the semantic gap influence.

References

[1] Marcelo CO, Raniery FJ. A Bag-of-Tasks Approach to Speed Up the Lung Nodules Retrieval in the BigData age. In Proceedings of 15th IEEE International Conference on e-Health Networking, Applications and Services 2013: 632 – 636.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1022

Ontology-Driven Semantic Search for Brazilian Portuguese Clinical Notes

Sadid A. Hasan[a], Xianshu Zhu[a], Joey Liu[a], Claudia M. Barra[b], Lucas Oliveira[b], Oladimeji Farri[a]

[a] *Philips Research North America, New York, United States*
[b] *Pontifical Catholic University of Paraná, Curitiba, Brazil*

Abstract

The emerging penetration of Health IT in Latin America (especially in Brazil) has exacerbated the ever-increasing amount of Electronic Health Record (EHR) clinical free text documents.This imposes a workflow efficiency challenge on clinicians who need to synthesize such documents during the typically time-constrained patient care. We propose an ontology-driven semantic search framework that effectively supports clinicians' information synthesis at the point of care.

Keywords:

Natural Language Processing (NLP); Semantic Search; Ontology Classification; Brazilian Portuguese

Introduction

Traditional semantic search applications mainly focus on the general knowledge domain while understanding the English language. Furthermore, existing semantic search applications based on Brazilian Portuguese are not designed to support routine healthcare tasks and clinical workflow. We propose an ontology-driven semantic search framework to provide clinicians better support in their decision-making process.

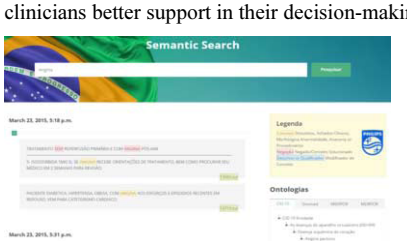


Figure 1 - Screenshot of semantic search user interface

Methods

Our semantic search framework uses open source NLP toolkits to analyze the Brazilian Portuguese grammar. The resulting grammatical units are processed for clinical concept extraction and mapping to Concept Unique Identifiers (CUI) in the Unified Medical Language System (UMLS) Metathesaurus. We generate relevant hierarchical representations of clinical concepts in the Portuguese version of the International Classification of Diseases (CID-10) based on publicly available language translations, and mapped these concepts to their English equivalents within the UMLS, including the Portuguese clinical ontologies: MSHPOR (Medical Subject Headings in Portuguese) and MDRPOR (Medical Dictionary for Regulatory activities in Portuguese). CID-10 concepts are then mapped to the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). Our ability to achieve Portuguese-English mapping of controlled clinical vocabularies represents a significant contribution to Brazilian Portuguese clinical NLP research. Our framework also offers patient data visualization via an intuitive search engine interface (Figure 1) that allows clinician users to efficiently retrieve relevant clinical information from large data sources.

Results

To evaluate the effectiveness of our semantic search framework, we conducted a pilot study using a subset of 80 cardiology discharge summaries written in Brazilian Portuguese. 4 clinical students (2 Medical and 2 of a Nursing background) were each asked to perform 2 clinical summarization tasks using a traditional EHR-like system (without semantic search functionality) for one task and our semantic search system for the second task. Our study results showed that on average, the students required 39.9% less time to complete task 1 and 10.7% less time to complete task 2 when they used our semantic search application. Further analysis of the results suggested that the differences in time to task completion may be due to each student's decision to read every discharge summary while using the traditional EHR-like system. In addition, qualitative data from the interviews conducted after task completion revealed that the students unanimously preferred having access to the semantic search application for task completion as they could search for keywords and browse through fewer summaries in contrast to manually synthesizing information. Limitations in our study include possible disparities in domain expertise, documentation speed, degree to which the tasks were understood by the students, and differences in task complexity. In our future research, we would perform a larger and more robust evaluation of our application while minimizing the aforementioned limitations.

Conclusion

Preliminary evaluation of our ontology-driven semantic search framework for Brazilian Portuguese clinical documents reveals that it effectively improves clinical task workflow and can contribute to the decision making process of the clinicians through faster, and more accurate information retrieval and visualization at the point of care.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1023

Use of Self-Service Query Tools Varies by Experience and Research Knowledge

Gregory W. Hruby[a], Jessica Ancker[b], Chunhua Weng[a]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*
[b] *Department of Biomedical Informatics, Cornel University, New York, NY, USA*

Abstract

The lack of understanding of user experience with self-service query tools is a barrier to designing effective query tools and is what propelled this study. User actions were documented and transformed into networks of actions for qualitative analysis. Proficient use of self-service query tools requires significant technical experience. To decrease the user learning curve, additional user education is necessary for novice users.

Keywords:
Query Tools; Expert and Novice Users.

Introduction

A major challenge to clinical and translational research resides in facilitating access to EHR data for researchers [1]. Self-service query tools (SSQT) have been developed to meet this need among diverse users [2]. Understanding how diverse users interact with SSQTs can inform designs for effective query tools in the future. Therefore, this study reports user experiences across SSQTs and research knowledge.

Methods

Eight semi-structured interviews and user observations were performed at four academic institutions. Users (physicians, clinical researchers, EHR data analysts) were divided into two sub-groups, experts and novices, based on having greater than 2 years of experience with research and the SSQT. Users were asked to perform a query to resolve their real-world information need using think-aloud protocols. All observations were videotaped to capture the actions and thoughts of the user. When the user completed the query, an exit interview was performed. A user-action schema was iteratively developed and pruned for video annotation. For each video, user actions were annotated with this schema by a single annotator, and then repeated for quality control. For each expert and novice group across research knowledge and SSQT experience, normalized directed network graphs were produced and user-action tables were created.

Results

We identified four user actions: browse, enter, review, and select, which had three, four, five, and nineteen subtasks, respectively. As shown in Table 1, expert and novice users had similar frequency distributions among actions. Expert researchers extensively used the action "Enter 'Search Criteria,'" while novice researchers rarely did.

Table 1 – User-Actions by SSQT experience and research knowledge

User-action	Research Knowledge		SSQT Experience	
	Expert (n=5)	Novice (n=3)	Expert (n=3)	Novice (n=5)
Browse	7	11	8	9
Enter	10	9	12	8
Review	4	5	4	4
Select	33	54	42	40

The directed network graphs produced show patterns of user-actions within our groups. Most notably, SSQT experts exhibited a more organized flow of user actions. They added data elements rather than removed them after reviewing the query build, and also tended to reformulate their queries after obtaining a result set. SSQT novices seemed to augment their queries by adding or removing data elements after reviewing their query build. Interestingly, research knowledge did not display varying patterns of user-actions. Both expert and novice researchers used the SSQT in a similar pattern.

Conclusion

As expected, SSQT experts seem to be more efficient with their actions when completing a query, implying that an improved user experience may be related to user education of the SSQT functionalities and uses. Additionally, SSQT experts seem to augment their query more often after reviewing query results. Finally, there seemed to be minimal differences between the user-action pattern of research experts and novices. However, research experts frequently performed data element searches rather than browsing.

Acknowledgments

This research was funded under NLM grant R01LM009886, and 5T15LM007079, and CTSA award UL1 TR000040.

References

[1] Greim, J., et al., The quality data warehouse: delivering answers on demand. AMIA Annu Symp Proc, 2006: p. 934.

[2] Deshmukh, V., S. Meystre, and J. Mitchell, Evaluating the informatics for integrating biology and the bedside system for clinical research. BMC medical research methodology, 2009. **9**(1): p. 70.

Address for correspondence

Gregory W. Hruby, Email: gh2162@columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1024

Comparision and analysis of top 10 exercise android Apps in mainland China

Yanling Wang[a], Liu Sun[a] , Yahong Xu[a] , Qian Xiao[a], Polun Chang[b], Ying Wu[a1]

[a] *School of Nursing, Capital Medical University, Beijing, China*
[b] *Institute of biomedical informatics, National Yang-Ming University, Taipei*

Abstract

Medical guidelines highly recommend physical activity and aerobic exercise in the prevention of primary and secondary cardiovascular disease. The use of exercise-promoting application software may improve clinical outcomes for cardiovascular disease (CVD) patients. The study aimed to compare and analyze the functions of the top 10 exercise Android Apps which had more than 1,000,000 downloads from the main four Android App stores in mainland China. The results showed that most of these popular apps had pedometer, exercise plan preset, user data presentation, user encouragement and community sharing functions while a few of them had exercise video clips or animation support and wearable devices. Given these data, the conclusion is that these popular apps fulfill some of the functions recommended by medical guidelines, however, lack of some functions such as pre-exercise risk assessment, the exercise intensity recording, specific instructions by professionals, and monitoring functions for CVD patients.

Keywords:
Exercise; Android; App; Health;CVD

Introduction

It is recommended that CVD patients engage in 3 to 4, 40 minute sessions per week of moderate-to-vigorous physical activity[1]. However, most patients do not satisfy these recommendations. It is proposed that the use of exercise apps by CVD patients may improve adherence to an exercise schedule, but it is unclear whether popular apps have the appropriate functions to support the CVD patient in fulfilling recommended physical activity guidelines.

Materials and Methods

The top 10 exercise apps were selected from Android App stores, which include Tencent, 360, Baidu, and 91 in China, all of which had more than 1000,000 downloads in the main Android App stores (by the time 19th Dec 2014). The data were stored and analyzed using Excel software.

Results

The general information of the top 10 android Apps is shown in Table 1. The main funtions of these apps are illustrated in Figue 1. Most of the top 10 apps have pedometer, exercise plan preset by users, user data presentation, user encouragement functions while a few of them had exercise video clip or anination support and wearable devices.

Conclusion

The top 10 Android Apps have some functions that can be used to assist CVD patients in meeting recommended physical activity guidelines, however, there is still a need to develop some special functions for patients with CVD.

Table 1– General Information on the top 10 Android Apps

App Names	Ver	Size (Mb)	Downloads（10,000） Tencent	360	Baidu	91
Sythealth	3.6.2	9.68	253	331	4280	>50
Gudong	5.2.0	17.81	114	269	320	>100
LedongV	3.3	18.81	95	178	241	>50
DailyYoga	6.0	6.96	66	427	283	>50
Pacer	2.5.1	5.84	162	39	33	>5
NikeRuning	1.5.1	33.84	25	78	86	>10
YidongGPS	2.4.0	7.05	21	23	473	>5
YueDong	2.1.0	9.28	49	69	659	>10
FitnessBook	2.2.7	13.32	49	424	86	>10
Push Ups	3.161	5.40	16	115	67	>10

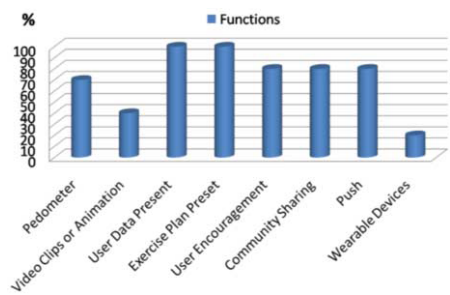


Figure 1–The main functions of the top 10 android apps

References

[1] Eckel RH, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014 Jun 24;129(25 Suppl 2):S76-99.

Address for correspondence

Ying Wu, RN, PhD, Tel.: 8610-8391-1766; fax: 8610-8391-1641. E-mail: helenywu@vip.163.com

[1] Corresponding author

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1025

3D Printed Models and Navigation for Skull Base Surgery: Case Report and Virtual Validation

Lucas E. Ritacco[1], Federico Di Lella[2], Axel Mancino[1], Fernan Gonzalez Bernaldo de Quiros[1], Carlos Boccio[2], Federico E. Milano[3,4]

[1]Virtual Planning and Navigation Unit, Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina, [2]ENT department, Hospital Italiano de Buenos Aires, Argentina. [3]Virtual Planning and Navigation Unit, Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina. [4]National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina.

Abstract

In recent years, computer-assisted surgery tools have become more versatile. Having access to a 3D printed model expands the possibility for surgeons to practice with the particular anatomy of a patient before surgery and improve their skills. Optical navigation is capable of guiding a surgeon according to a previously defined plan. These methods improve accuracy and safety at the moment of executing the operation. We intend to carry on a validation process for computed-assisted tools. The aim of this project is to propose a comparative validation method to enable physicians to evaluate differences between a virtual planned approach trajectory and a real executed course. Summarily, this project is focused on decoding data in order to obtain numerical values so as to establish the quality of surgical procedures.

Keywords:
Virtual preoperative planning; 3D printed model; skull base navigation.

Introduction

This study aims to show in a virtual preoperative plan setting, a surgical trajectory through the temporal bone. Afterwards, a printed 3D model is used in order to execute and practice the planning under navigated guideline. Finally, the same planned trajectory is executed in the patient, also under navigated guideline. The objective of this work is to establish a comparative virtual study for testing the results between the 3D planned trajectory and the executed trajectory.

Materials and Methods

A 38 years old man complains about sudden sensorineural hearing loss on the left side associated with tinnitus and transient peripheral facial paralysis in relation to physical efforts. With the diagnosis of cholesterol granuloma, an infracochlear approach was planned for apply a transtemporal tumor drainage assisted with intraoperative navigation.
A plastic 3D model was printed (Figure 1).
First, planning was executed in a 3D model under navigation guidance (Intellect navigation software; Stryker Navigator, Freiburg, Germany). Surgery was simulated by drilling the 3D model [1]. This experiment was carried on using surgical microscope simulating the setting of the operating room.

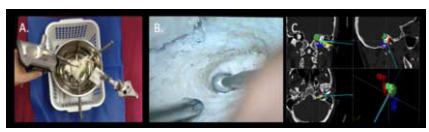


Figure 1- A. Navigating 3D plastic model. B. Drilling plastic model guided by navigation. C. Navigator showed the trajectory planned during the procedure to the surgeon.

Next, the drilled 3D model was CT scanned. Finally, these same steps were applied to the clinical case.
The trajectory discrepancies were measured computing the mean relative euclidean distance from the center line of the planned trajectory to the center line of the executed trajectory.

Results

Differences between the planned and the executed trajectories in the printed model were on average 2.62mm (SD 0.94mm). Differences between trajectory planned and and executed in the post-operative patient model were on average 0.69mm (SD 0.44mm).

Conclusion

We could state that 3D models and surgical navigation tools potentially improve quality and safety for interventions, since surgeons have an intraoperative three-dimensional orientation. It is a major advantage for physicians to work with computer-aided surgery tools since the results of the procedures could be evaluated using the proposed validation model.

References

[1] Narayanan V, Narayanan P, Rajagopalan R, Karuppiah R, Rahman ZAA, Wormald P-J, et al. Endoscopic skull base training using 3D printed models with pre-existing pathology. Eur Arch Otorhinolaryngol [Internet]. 2014 Oct 8; Available from: http://dx.doi.org/10.1007/s00405-014-3300-3

Address for correspondence

Juan D. Peron 4190, C1181ACH Capital Federal, Buenos Aires.
Virtual Planning and Navigation Unit, Department of
Health Informatics, Hospital Italiano de Buenos Aires, Argentina.
lucas.ritacco@hiba.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1026

Accuracy of Chest Wall Tumor Resection Guided by Navigation: Experimental Model

Lucas E. Ritacco[a], David E. Smith[b], Axel V. Mancino[a], German L. Farfalli[c], Luis Alberto Aponte-Tinao[c], Federico E. Milano[a,d]

[a] *Virtual Planning and Navigation Unit, Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*
[b] *Department of General Surgery, Hospital Italiano de Buenos Aires, Argentina*
[c] *Department of Orthopedic Surgery – Orthopedic Oncology, Hospital Italiano de Buenos Aires, Argentina*
[d] *National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina*

Abstract

Difficulty in identification wall chest tumors lead to unnecessary wide resections. Optical navigation and preoperative virtual planning are assets for surgeries that require exactness and accuracy. These tools enable physicians to study real anatomy before surgery and to follow an established pathway during procedure ensuring effectiveness. The aim of this paper is to demonstrate that Preoperative Virtual Planning is a useful tool in chest tumor interventions to define oncological margins successfully. Moreover, it is possible to use a virtual specimen in order to quantify accuracy. Optical navigation has been used in surgical procedures such as neurosurgery, orthopaedics and ENT over the last ten years. This principle is used in order to orientate the surgeon in three dimensional spaces during the surgery. Surgeons are guided intraoperatively with navigation and are able to obtain a correspondence between images acquired and processed before the surgery and the real anatomy.

Keywords:

Optical navigation; Chest wall tumor; Preoperative virtual planning; Chest wall allograft**.**

Introduction

The main problem of chest wall tumors is the non-visible or non-palpable oncology margins that lead to wide surgical resections. Preoperative planning and navigation introduce a new concept in computer-assisted surgery that consists of using specific tools, which provide safe and accurate localization to optimize the surgical margins while keeping the safety margins unharmed [1]. The objective of this paper is to describe technical aspects and to establish an experimental design in order to demonstrate accuracy in a chest wall navigation.

Methods

The four cases included in this work were patients with a chest wall tumor where the margin cannot be determined using direct visual inspection nor palpable features. A navigated pointer was used to mark the osteotomy, which had been previously planned, on the superficial bone with a surgical marking pen. Next, the surgeons performed the osteotomy following this mark with a freehand saw. After surgery, the surgical specimen was CT scanned, reconstructed and registered to the bone cortex in the plan. A plane fitting algorithm was used to find the precision of the osteotomy (Fig 1).

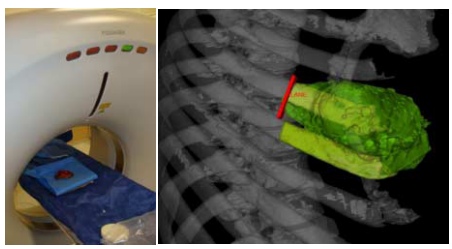


Figure 1 – The specimen is CT scanned and registered (in green) to the preoperative plan (in white) to measure the accuracy and precision.

Results

The accuracy (median) of the osteotomy was of 0.44mm with a precision (equivalent to a 6 sigma dispersion) of 3.78mm. The 99.99966% of the osteotomy surface is below the tolerance threshold of 2mm. In the four cases, it was possible to generate a preoperative plan and achieve a good intraoperative registration between the virtual scenario and the patient.

Conclusion

Navigation is an essential computer-assisted tool for surgery, which should be taken into account in order to assist the surgeon in the operating room. This technique is accurate and reproducible. Furthermore, this tool shows exactness and potentially reduces morbidity during surgery. The present validation model, using a virtual specimen scanned, is an useful and novel method to quantify results after a chest wall tumor resection is executed.

References

[1] Wadley J, Dorward N, Kitchen N, Thomas D. Pre-operative planning and intra-operative guidance in modern neurosurgery: a review of 300 cases. Ann R Coll Surg Engl. Royal College of Surgeons of England 1999: Jul 81(4):217.

Address for correspondence

Hospital Italiano de Buenos Aires.Juan D. Peron 4190, C1181ACH Capital Federal, Buenos Aires. lucas.ritacco@hiba.org.ar

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1027

Identification of Incidental Pulmonary Nodules in Free-text Radiology Reports: An Initial Investigation

Lucas Oliveira[a], Ranjith Tellis[a], Yuechen Qian[a], Karen Trovato[a] and Gabe Mankovich[a]

[a] Clinical Informatics Solutions and Service, Philips Research North America, NY, USA

Abstract

Advances in image quality produced by computed tomography (CT) and the growth in the number of image studies currently performed has made the management of incidental pulmonary nodules (IPNs) a challenging task. This research aims to identify IPNs in radiology reports of chest and abdominal CT by Natural Language Processing techiniques to recognize IPN in sentences of radiology reports. Our preliminary analysis indicates vastly different pulmonary incidental findings rates for two different patient groups.

Keywords:

Incidental Findings; Natural Language Processing; Radiology; Patient Care; Algorithms.

Introduction

Due to increasing volume and quality of CT imaging studies, more incidental findings (IFs) are reported by radiology departments [1,2]. An incidental finding is an asymptomatic lesion noted in an imaging study that is unrelated to the 'reason for exam' [1,2,3]. Incidental pulmonary nodules (IPNs) are not unusual IFs for patients presenting with abdominal pain, for example.

IFs are communicated through the free-text narrative radiology report. This makes automated identification and longitudinal follow-up challenging, leading to untimely diagnosis or care delivery. A tool to accurately identify IFs can help the healthcare enterprise assess appropriate recommendations and track follow-ups in their patients. This work describes an initial step to identify IPNs in radiology reports of chest and abdomen CTs by applying natural language processing (NLP) techniques to recognize and quantify IPNs in sentences from radiology reports.

Methods

This study was performed on a sample of 3,785 representative, de-identified chest and abdomen CT reports from an academic hospital in Illinois, US**.** The reports were parsed with a sentence and section detection engine that also normalizes section headers (e.g., 'Clinical History', 'Findings', 'Impressions'). IPNs were detected using one list (L_1) with pulmonary nodule keywords (e.g., mass, lesion, or nodularity) and another list (L_2) with expressions indicating that a pulmonary nodule is not a new finding (e.g., again noted, stable, or previously). Regular expressions were created based on L_1 and L_2 to establish if a word or phrase was contained in a sentence. An IPN is defined as a match of L_1 in the 'Findings' section whose sentence does not match L_2 and the'Reason for Exam' and 'Clinical History' sections had no matches of L_1 indicating any prior pulmonary condition.

For a initial evaluation of the precision of our algorithm, we randomly selected 100 reports for manual review by one of the authors (KT): 50 reports with an IFs identified and 50 reports with no IFs identified.

Results

The results obtained in our preliminary investigation suggested that 9.2% (347/3,785) of reports contained at least one IPN mention. The rate of false positives and false negatives detected in the manual validation were 12% (12/100) and 2% (2/100) respectively. It has been reported [2,3,4] that the rate of IPNs are associated with risk factors such as age, tabacco use and cancer history. The rate of IPNs in reports for oncology patients (i.e., reports containing "metastasis" or "cancer" in *Reason for Exam* and *Clinical History* sections) was 10.8% (276/2,565 reports) compared to 5.8% (71/1,220) for non-oncology patients.

Conclusion

We have an initial NLP method for determining IPNs and we were able to successfully run it on a sizeable corpus of reports producing a rate of IPNs that is different than earlier rates discussed in the literature [2,3.4]. A more specific definition of IFs might be considered to exclude oncology patients. Next steps include further validation with clinical experts and a more expanded statistical evaluation will help assess the true clinical potential of our methods.

References

[1] MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett J.H, Naidich D.P, Patz E.D, Swensen S.J. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. Radiology 2005: 237 (2):295-400.

[2] Lumbreras B, Donat L, Hernández-Aguado I. Incidental findings in imaging diagnostic tests: a systematic review. Br J Radiol. 2010 : 83(988):276-289.

[3] Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'Byrne MM, Port JD, Tremaine WJ, Pitot HC, McFarland EG, Robinson ME, Koenig BA, King BF, Wolf SM. ncidental Findings in Imaging Research: Evaluating Incidence, Benefit, and Burden. Arch Intern Med 2010:170(17):1525-1532.

[4] Blagev DP, Lloyd JF, Conner K, Dickerson J, Adams D, Stevens SM, Woller SC, Evans RS, Elliott CG. Follow-up of incidental pulmonary nodules and the radiology report. *J Am Coll Radiol* 2014:11(4):378-283.

Address for correspondence

Lucas Oliveira – lucas.oliveira@philips.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1028

Follow-up Recommendation Detection on Radiology Reports with Incidental Pulmonary Nodules

Lucas Oliveira[a], Ranjith Tellis[a], Yuechen Qian[a], Karen Trovato[a] and Gabe Mankovich[a]

[a] *Clinical Informatics Solutions and Service, Philips Research North America, NY, USA*

Abstract

The management of follow-up recommendations is fundamental for the appropriate care of patients with incidental pulmonary findings. The lack of communication of these important findings can result in important actionable information being lost in healthcare provider electronic documents. This study aims to analyze follow-up recommendations in radiology reports containing pulmonary incidental findings by using Natural Language Processing and Regular Expressions. Our evaluation highlights the different follow-up recommendation rates for oncology and non-oncology patient cohorts. The results reveal the need for a context-sensitive approach to tracking different patient cohorts in an enterprise -wide assessment.

Keywords:

Radiology, Incidental Findings, Follow-up Recommendation, Natural Language Processing, Patient Care.

Introduction

Incidental findings (IFs) in radiology reports are image observations that are not directly related to the original aims of the imaging examination [1,2]. Attentive management of IFs after identification may lead to early diagnosis and treatment of disease. Often when IFs are communicated in the radiology report, guideline-specific follow-up recommendations, such as the Fleischner Guidelines [1], are lacking. Clear communication of follow-up recommendations by the radiologist will contribute to timely follow-up of IFs, which may improve patient outcomes, reduce healthcare costs, minimize patient radiation exposure, and avoid costs incurred by litigation for missing IFs [1,2,3]. This work describes an investigation of the IPN follow-up recommendation rate in free-text radiology reports. We present a methodology to automatically identify follow-up recommendations of IPNs in chest and abdomen computed tomography (CT) reports through the use of natural language processing. We then use it to identify follow-up rates for two patient cohorts.

Methods

This study was performed on the complete corpus of 3,785 de-identified chest and abdomen CT reports from an academic hospital in Illinois, US. The reports were first parsed into a sentence, paragraph and sections. The section detector normalizes headers to common sections (e.g., 'Clinical History', 'Findings' and 'Impressions'). Incidental pulmonary nodules were identified in 347 reports using an NLP engine whose internal working/evaluation is not in the scope of this paper.

The follow-up recommendations were detected using a list of commonly used expressions such as 'may be performed', 'follow-up', and 'could be considered'. Regular expressions were created based on this list to identify recommendations in paragraphs from the 'Impression' section.

Patients were categorized into two cohorts: oncology and non-oncology. Oncology patients are identified by phraseology in the 'Reason for Exam' or 'Clinical History' sections having words related to metastasis or cancer, but not lung cancers.

Results

In this evaluation, 9.2% (32/347) of the incidental pulmonary nodules had follow-up recommendations. Oncology patients had a 6.2% (17/276) follow-up recommendation rate, whereas non-oncology patients had a 21% (12/71) recommendation rate.

Conclusion

This study presents a methodology that identifies follow-up recommendation rates in radiologyreports. Over 91.8 % of reports containing pulmonary incidental findings failed to provide a recommendation. For oncology patients, the recommendation rate is especially low, possibly due to comorbid oncological conditions that already require regular follow-up. Our results demonstrate a different recommendation rate for two cohorts of patients. Next steps include further validation of the detection algorithm using multiple radiologist ground-truth annotations.

References

[1] MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett J.H, Naidich D.P, Patz E.D, Swensen S.J. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. Radiology 2005: 237 (2):295-400.

[2] Blagev DP, Lloyd JF, Conner K, Dickerson J, Adams D, Stevens SM, Woller SC, Evans RS, Elliott CG. Follow-up of incidental pulmonary nodules and the radiology report. *J Am Coll Radiol* 2014:11(4):378-283.

[3] Masciocchi M, Wagner B, Lloyd B. Quality review: Fleischner criteria adherence by radiologists in a large community hospital. *J Am Coll Radiol.* 2012: 9(5):336-339

Address for correspondence

Lucas Oliveira – lucas.oliveira@philips.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1029

Clinical Trial Feasibility Study Questionnaire Analysis

Iñaki Soto-Rey[a], Martin Dugas[a], Fleur Fritz[a]

[a] *Institute of Medical Informatics, University of Münster, Germany*

Abstract

With the growing complexity and cost of clinical trials (CTs) over the past few decades, Protocol Feasibility (PF) studies have become one of the most critical CT steps in order to avoid costly protocol amendments and ensure the success of CTs. The PF process includes interaction with clinicians located at targeted clinical sites, which results in slow and cumbersome process steps. These process steps are normally supported by information systems that allow users to create, share and collect responses to feasibility questionnaires. This investigation analyzes the systems and questionnaires utilized at several clinical research companies for PF. In addition, it provides recommendations that could eventually improve current methods and systems in place.

Keywords:

Clinical Trials as Topic, Feasibility Studies, Questionnaires.

Introduction

Clinical trials (CTs) often suffer delays, and their initial budget is adjusted upwards due to both recruitment rates not being met, and costly protocol amendments. A good trial protocol feasibility (PF) study has been proven to be an effective method to avoid such issues.[1]. The design of PF studies is currently supported by clinicians located at the targeted sites, often resulting in slow and cumbersome process steps that involve a great amount of resources and time [2]. This research analyzes common processes, information systems, and feasibility questionnaires (FQs), and advises how they could be improved.

Methods

The European Federation of Pharmaceutical Industries and Associations (EFPIA) companies participating in the EHR4CR project[1] were asked to deliver examples of FQs. Furthermore, we analyzed how they are designed and which process steps are required to obtain their responses. The received FQs and templates were manually reviewed and the following information was extracted: Name and number of sections, question types, and number of questions per section.

Results

Feasibility experts from seven EFPIA companies collaborated in this study. Seven processes, five PF questionnaire templates, and sixteen FQ examples were analysed.

One of the companies uses a word processor to build the FQ, three corporative systems and three free-to-use online survey systems. The number of FQ sections varies between 3 and 13 with one to 47 questions per section and 22 to 100 questions per questionnaire. In Table 1, the question types and number of questions are presented.

Table 1– Question types and number (N) of questions (Q)

Q type	Total QN	Q per template	Q per questionnaire
Free text	532	21,2	26,6
Radio Button	299	13,6	14,4
Radio Button +Comment	251	12	11,9
Number	230	4	13,1
Checkbox +Comment	62	4,8	2,4
Checkbox	14	0,2	0,7
Free text table	12	0,4	0,8
Date	6	0	0,4

Discussion

This research identified a total of eight different question types in FQs, of which free text was the most common one. Most of the companies use templates or pre-defined questions to build the FQs but – even though the questions are frequently repeated – the analysed systems do not use historical records to allow auto-completing answering. Interviewees reported the number of PF design tools and the long-time interval to obtain the FQ responses as the most critical PF design issues.

The use of free text questions should be reduced due to the intricacy of free text answer completion and analysis. Electronic survey systems must be improved with the re-use of historical data to auto-complete responses, which facilitates responding and speeds up the entire PF process.

Conclusion

Free text questions in feasibility questionnaires need to be avoided and PF survey systems need to re-use data to automatize responses. There is need for a system that congregates all features of PF design and streamlines the methods in place.

References

[1] K. Getz, “Improving protocol design feasibility to drive drug development economics and performance,” *Int. J. Environ. Res. Public. Health*, vol. 11, no. 5, pp. 5069–5080, May 2014.

[2] I. Soto-Rey, B. Trinczek, T. Karakoyun, M. Dugas, and F. Fritz, “Protocol Feasibility Workflow Using an Automated Multi-country Patient Cohort System,” *Stud. Health Technol. Inform.*, vol. 205, pp. 985–989, 2014.

Address for correspondence

Iñaki Soto Rey – Inaki.SotoRey@uni-muenster.de

[1] http://www.ehr4cr.eu/

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1030

Extraction Of Adverse Events From Clinical Documents To Support Decision Making Using Semantic Preprocessing

Jan Gaebel[a], Till Kolter[b], Felix Arlt[c], Kerstin Denecke[a]

[a] *Innovation Center Computer Assisted Surgery, Leipzig, Germany*
[b] *ID Information und Dokumentation im Gesundheitswesen, Berlin, Germany*
[c] *Clinic for Neurosurgery, University Hospital Leipzig, Leipzig, Germany*

Abstract

Clinical documentation is usually stored in unstructured format in electronic health records (EHR). Processing the information is inconvenient and time consuming and should be enhanced by computer systems. In this paper, a rule-based method is introduced that identifies adverse events documented in the EHR that occurred during treatment. For this purpose, clinical documents are transformed into a semantic structure from which adverse events are extracted. The method is evaluated in a user study with neurosurgeons. In comparison to a bag of word classification using support vector machines, our approach achieved comparably good results of 65% recall and 78% precision. In conclusion, the rule-based method generates promising results that can support physicians' decision making. Because of the structured format the data can be reused for other purposes as well.

Keywords:

Medical Language Processing; Information Extraction; Electronic Health Records; Drug-Related Side Effects and Adverse Reactions; Clinical Decision Support Systems.

Introduction

Clinical documentation serves as a information source for medical professionals. However, identifying and processing the relevant information is time consuming [1]. To support information retrieval, clinical decision support systems (CDSS) are developed. One major challenge is, that the majority of the data is stored in unstructured, narrative format. Natural language processing (NLP) methods offer a solution for identifying text passages that are relevant to a specific question or task.

In this paper, we introduce a rule-based NLP method for identifying adverse events in clinical documents. These events influence the following treatment and health care planning and thus need to be considered by a physician.

Methods & Materials

In the course of a manual analysis of clinical documents in German with respect to adverse events, we recognized that there are no significant syntactical similarities in the wording of adverse events. For this reason, we decided to chose a semantic approach instead of a syntactical apporach for adverse event extraction. Our approach exploits the terminology server ID MACS® for normalizing and structuring texts semantically. Medical and linguistic concepts are identified. If a medical concept is found, its severity is assessed by analyzing its surrounding linguistic concepts. When certain words express negativity, the passage is stored as an adverse event.

Evaluation and Results

We compared the results of our method with those of a standard machine learning classifier implemented in WEKA [2] using a test dataset of five clinical documents from the neurosurgical department that were annotated by physicians. First, we performed a 10-fold-cross validation with a Support Vector Machines algorithm applied on a bag of words representation of the text. Second, we exploited this trained model to the sentences of the test dataset. The sentences of the test set were also classified using our approach. Precision and recall are shown in Table 1.

Table 1 – Recall, precision, fallout, F-measure

	WEKA	AdverseEvent-Extraction
Recall	0,7647	0,6470
Precision	0,4642	0,7857
F-measure	0,5777	0,7096

Conclusion

We conclude, that an extraction of complex linguistic entities is possible, but not with plain syntactical methods, like assessing word frequencies or regular expressions. A semantic preprocessing of the text and usage of medical terminologies are necessary to achieve good results. Thereby, information extraction systems can combine specific, standardised concepts to complex items or events. The output of such systems includes a semantically structured and normalised representation of the desired information. In this way, the extraction result can be processed further. With our method, the results can be used for calculations as for instance in decision support systems, statistical analysis or administrative calculations.

References

[1] Denecke K, Bernauer J. Extracting Specific Medical Data Using Semantic Structures. AIM: 4594; pp. 257-264.

[2] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The weka data mining software: An update. SIGKDD Explor. Newsl., 2009: 11(1); pp. 10–18.

Address for correspondence

Jan Gaebel
University Leipzig, Faculty of Medicine
Innovation Center Computer Assisted Surgery (ICCAS)
Semmelweisstr. 14, 04103 Leipzig, Germany

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1031

Development and evaluation of task-specific NLP framework in China

Caixia Ge[a], Yinsheng Zhang[a], Zhenzhen Huang[a], Zheng Jia[a], Meizhi Ju[a], Huilong Duan[a] and Haomin Li[b,c]

[a] *College of Biomedical Engineering and Instrument Science, Zhejiang University, The Key Laboratory of Biomedical Engineering, Ministry of Education, Hangzhou, China*
[b] *The Children's Hospital of Zhejiang University School of Medicine, Hangzhou, China*
[c] *Institute of Translational Medicine, Zhejiang University, Hangzhou, China*

Abstract

Natural language processing (NLP) has been designed to convert narrative text into structured data. Although some general NLP architectures have been developed, a task-specific NLP framework to facilitate the effective use of data is still a challenge in lexical resource limited regions, such as China. The purpose of this study is to design and develop a task-specific NLP framework to extract targeted information from particular documents by adopting dedicated algorithms on current limited lexical resources. In this framework, a shared and evolving ontology mechanism was designed. The result has shown that such a free text driven platform will accelerate the NLP technology acceptance in China.

Keywords:

Natural language processing; Information extraction; Clinical task.

Introduction

Directly using data recorded in Electronic Medical Record (EMR) systems for diverse clinical tasks will increase the automation of information systems in healthcare. However, the data formats and structure required by different tasks are seldom fulfilled by most EMRs, especially in Chinese hospitals where most clinical information is recorded in clinical documents using natural language. In this study, we propose a task-specific Natural Lanugage Processing (NLP) framework which customizes algorithms and defines targeted concepts and clinical documents for diverse information extraction tasks. This NLP framework seamlessly integrates with a general CDS framework by sharing a common ontology and provides an evolving approach to accumulating lexical resources.

Methods

Since 2012, we designed and developed a general CDS framework in a 2000-bed hospital. As reported in many studies, many CDS applications that require additional data entry have very poor acceptance in the clinical setting. So, an NLP framework was proposed to automatically extract information and facilitate CDS data entry. In this setting, a homegrown medical ontology that contains clinical concepts, and relationships between concepts, will be used in the CDS platform and is maintained through a web portal. As most information extraction tasks only focus on specific concepts, a task-specific NLP framework with a lexicon self-propagation mechanism was proposed. As shown in Figure 1, by defining NLP tasks to customize NLP algorithms, and targeting document types and related concepts, task-specific concept sets will be generated and used to complement the homegrown ontology. This will make the lexicon meet the requirements of different tasks. Through integration with EMRs, a task monitor was developed to generate tasks when certain types of clinical documents were released. Different NLP task instances will be invoked based on the defined task profiles. In each NLP task instance, targeted concepts will be extracted from specific documents and fed to the platform for the targeted task.

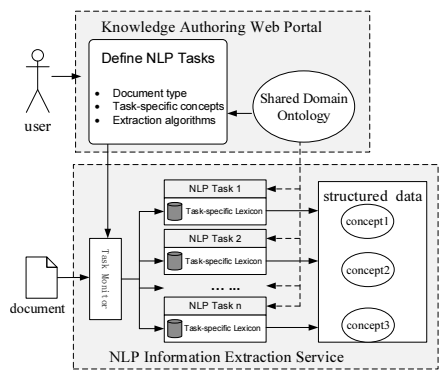


Figure 1- Task specific NLP framework

Users can define new NLP profiles for different tasks. A plug-in mechanism based on the .Net Framework Reflection technology was developed to allow this framework to load and execute third-party NLP algorithms at runtime.

Results

To evaluate this NLP framework we used three different information extraction tasks. In task I, a concept-value pair extraction task that extracts the measurement value of specific concepts from obstetric ultrasound reports was defined. The second task identified adverse drug events (ADE) from progress notes. The last task was to extract general purpose symptoms from progress notes to build a symptom timeline for each patient. The evaluation results are shown in Table 1.

Table 1 - Evaluation Results

	Task I	**Task II**	**Task III**
Recall	97.8%	NA	NA
Precision	98.5%	80.8%	98.2%

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1032

Extracting Dependence Relations from Unstructured Medical Text

Charles Jochim [a], Yassine Lassoued [b], Bogdan Sacaleanu [a], Léa A. Deleris [a]

[a] *IBM Research – Ireland, Dublin, Ireland*
[b] *CMRC – University College Cork, Cork, Ireland*

Abstract

Dependence relations among disease and risk factors are a key ingredient in risk modeling and decision support models. Currently such information is either provided by experts (costly and time consuming) or extracted from data (if available). The published medical literature represents a promising source of such knowledge; however its manual processing is practically infeasible. While a number of solutions have been introduced to add structure to biomedical literature, none adequately recover dependence relations. The objective of our research is to build such an automatic dependence extraction solution, based on a sequence of natural language processing steps, which take as input a set of MEDLINE abstracts and provide as output a list of structured dependence statements. This paper presents a hybrid pipeline approach, a combination of rule-based and machine learning algorithms. We found that this approach outperforms a strictly rule-based approach.

Keywords:

Risk; Risk Factors; Artificial Intelligence; Natural Language Processing; Data Mining.

Introduction

Bayesian belief networks are a convenient tool for macro-level risk models, i.e., risk models articulating the dependence between multiple risk factors, diseases and conditions. They have long been advocated as a useful decision and risk modeling framework in medicine [1]. Our focus on automating the extraction and aggregation of risk information seeks to address the practical challenge of building such networks. We focus specifically on the information contained in MEDLINE, which has been steadily growing at a rate of about 1M publications per year in the past few years and for which manual consumption is no longer practical. The larger context of this research is to build a system, called Medical Recap, which automatically extracts risk information from medical papers and then aggregates this knowledge into a Bayesian network.

Methods

This paper focuses on the sub task of extracting dependence information. From the sentence "Smoking increases the risk of lung cancer", we want to extract the relation (Smoking, Lung Cancer). Our approach to this challenge can be divided into two sub-problems: entity detection and relation construction. Entity detection deals with the identification of the elements in a sentence that are part of a relation. Relation construction is about articulating those elements together (which is not straightforward when multiple relations are expressed in the same sentence). We rely on machine learning for entity detection and propose a rule-based algorithm for defining the relations between the extracted entities. More specifically, we have developed a pipeline that identifies sentences containing dependence relations, then identifies the entities (i.e., dependence relation variables) in the sentence using conditional random fields (CRF) [2]. Dependence relations are constructed from the so-obtained entities using a set of rules, which accurately identify correct dependence relations when given the correct entities.

Results

We evaluated our pipeline both after entity detection and after relation construction. The performance of our proposed hybrid approach consistently outperformed a strictly rule-based baseline that uses shallow syntactic parsing. For entity detection, we achieved an F_1 score of 67.1 vs. the baseline performance of 62.6. Likewise, for relation construction using the entities we extracted, we reached an F_1 score of 53.6 vs. a baseline F_1 of 49.1. The results showed that our approach has higher precision and lower recall across the various steps of the pipeline, which is preferred for our particular use case.

In our error analysis of the entity detection we found a number of false negative errors where correct entities were not detected. Sometimes the CRF classifier also had difficulty detecting the correct entity boundaries. These problems were due in a large part to the heterogeneous nature of the entities and the modest size of our annotated corpus.

Conclusion

Our work investigates how to automatically extract dependence information from the academic medical literature, thereby streamlining an otherwise time-consuming and costly process. We developed a set of NLP steps to address this issue and found that our pipeline of algorithms performed satisfactorily especially compared to a naïve baseline.

Acknowledgments

This project was supported in part through a grant from the Irish Innovation Development Agency (ref. 133954, September 2009) and Science Foundation Ireland (13/IF/I26291).

References

[1] Pauker SG, Wong JB. The influence of influence diagrams in medicine. Decision Analysis 2005: 2: 238–244

[2] Lafferty JD, McCallum A, and Pereira FCN. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, 2001: 282–289.

Address for correspondence

Charles Jochim – charlesj@ie.ibm.co

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1033

A Frequency-based Strategy of Obtaining Sentences from Clinical Data Repository for Crowdsourcing

Dingcheng Li[a], Majid Rastegar Mojarad[a], Yanpeng Li[a], Sunghwan Sohn[a], Saeed Mehrabi[a], Ravikumar Komandur Elayavilli[a], Yue Yu[a,b], Hongfang Liu[a]

[a] *Mayo Clinic, Rochester, MN, USA*
[b] *Department of Biomedical Informatics, University of Jilin, Jilin, China*

Abstract

In clinical NLP, one major barrier to adopting crowdsourcing for NLP annotation is the issue of confidentiality for protected health information (PHI) in clinical narratives. In this paper, we investigated the use of a frequency-based approach to extract sentences without PHI. Our approach is based on the assumption that sentences appearing frequently tend to contain no PHI. Both manual and automatic evaluations on 500 sentences out of the 7.9 million sentences of frequencies higher than one show that no PHI can be found among them. The promising results provide potentials of releasing those sentences for obtaining sentence-level NLP annotations via crowdsourcing.

Keywords:

patient health information, clinical notes, high-frequent sentences, bigram filtering, crowdsourcing, de-identification

Introduction

Crowdsourcing has emerged as a popular method to generate training data for machine learning, natural language processing (NLP) and related fields[1,2]. However, in clinical NLP, no one has taken advantage of crowdsourcing since it may reveal the protected health information (PHI). Until now, no de-identification tools could guarantee 100% PHI removal from clinical narratives. Consequently, it remains infeasible to employ crowdsourcing for generating training data for clinical NLP. In this work, we propose a simple frequency-based approach to extract sentences containing no PHI under the assumption that sentences appearing frequently tend to contain no PHI. Based on this approach, it could be possible to generate training data through crowdsourcing for various sentence-level clinical NLP tasks such as concept identification or relation extraction.

Materials and Methods

In this work, we used all of electronic medical records (EMR) extracted from enterprise data trust (EDT) of Mayo Clinic. The EDT stores structured data and unstructured texts from a comprehensive snapshot of Mayo Clinic's service areas. It includes clinical notes, hospital summary, post-procedure notes, procedure note, progress note, tertiary trauma and transfer note. It is composed of 39.7 million records and 1.7 million tokens. We collected high-frequent sentences using following steps: building bigram repository; constructing sentence repository; bigram filtering; sentence frequency filtering and segmentation; dictionary-lookup filtering and sentence distribution analysis.

Experiments

System Results

The running of the above-described workflow yielded about 7.9 million unique sentences with frequencies higher than one (the total number was 276.7 million with an average frequency of 35). We then divided the whole sentence repository into 10 intervals based on log2 frequency. Although the numbers of sentences for each interval were different, we randomly sampled 50 sentences at each interval for evaluation. Meanwhile, we also sampled 200 more sentences the same way from sentences which had passed bigram filtering but only appeared once (the total number of such sentences is 109.2 million) for comparison purpose. Accordingly, the total number of sentences for evaluation is 700.

Manual Evaluation

The 700 sentences were assigned to four experienced reviewers to assess whether PHI could be found among them. In the event that any such information was found, reviewers were required to fill name, profession, location, age, date, contact, ids and comments for other sensitive information into a spreadsheet. The evaluation results (the union of the four reviewers' results), showed that, as expected, most PHI occurred in the first interval. The most frequent PHI element was name – appearing 7, 3, and 2 times in the first, second, and third interval respectively. Similar patterns were seen for profession as well. For age and date, 4 and 13 times were in the first interval and 1 and 3 times in the second interval. No contact was found for all intervals and ids showed one time in the first three intervals.

Conclusion

In summary, we proposed a method based on frequencies to extract sentences containing no PHI from a clinical data repository. We experimented in utilizing Mayo's EDT clinical notes. About 7 million unique sentences which appear more than two times were extracted. The final evaluation on 700 sampled sentences show that nearly no PHI can be found from sentences with higher frequencies (500 of them have frequencies higher than one). As follow-up steps, we will explore ways to make the crowdsourcing for clinical notes realistic and develop corresponding systems.

Acknowledgments

This work was made possible by joint funding from National Institute of Health R01GM102282, R01LM11369, R01LM11829, and R01 LM011934.

References

[1] W. Safire. (2009, Feb. 5) On Language. *New York Times Magazine*.

[2] H. Zhai, T. Lingren, L. Deleger, Q. Li, M. Kaiser, L. Stoutenborough*, et al.*, "Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing," *Journal of medical Internet research,* vol. 15, 2013.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1035

Extraction of Vital Signs from Clinical Notes

Olga V. Patterson, PhD[a,b], Makoto Jones, MD[a], Yiwen Yao, MS[a], Benjamin Viernes, MPH[a], Patrick R. Alba, BA[a,b], Theodore J. Iwashyna, MD, PhD[c], Scott L. DuVall, PhD[a,b]

[a] VA Salt Lake City Health Care System; [b] University of Utah, Salt Lake City, UT;
[c] Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, MI

Abstract

Assessment of vital signs is an essential part of surveillance of critically ill patients to detect condition changes and clinical deterioration. While most modern electronic medical records allow for vitals to be recorded in a structured format, the frequency and quality of what is electronically stored may differ from how often these measures are actually recorded. We created a tool that extracts blood pressure, heart rate, temperature, respiratory rate, blood oxygen saturation, and pain level from nursing and other clinical notes recorded in the course of inpatient care to supplement structured vital sign data.

Keywords:

Natural language processing; Vital signs.

Introduction

To support a study aimed at measuring the prevalence and incidence of sepsis, severe sepsis, and septic shock among veterans, we looked at the availability of vital signs and relevant laboratory data across the Department of Veterans Affairs (VA) medical centers. The VA electronic medical record includes a vital sign package, but we found in an initial assessment that the intensive care units (ICU) of 44 facilities and the acute medical care wards of 7 facilities had less than 80% of patient days where vitals were recorded. Given that it is mandatory for even non-intensive care units to record vitals at least once per shift, and the vital signs are usually summarized in patient care notes, we explored natural language processing (NLP) as a way of filling the gaps in the structured data.

Methods

The initial goal for the NLP system was to extract blood pressure, heart rate, and temperature values from a wide range of clinical notes. A random set of 5,000 documents was selected for a cohort of VA inpatient stays from 2000 to 2012. The documents were manually annotated with instances of vital sign terms and values to serve as a reference standard. The reference standard also included a section header annotation that indicated the beginning of vital signs section. Distinct strings of the vital sign terms were compiled into a lexicon expressed as a set of regular expressions. Analysis of the annotated set revealed a small number of patterns that linked terms and values. The NLP system was built based on the lexicon and patterns collected directly form the annotated corpus. The system utilizes Leo, a set of libraries that enable more efficient utilization of Unstructured Information Management Architecture Asynchronous Scaleout (UIMA AS)[1]. Document review also revealed that other types of vitals signs (weight, height, oxygen saturation, respiratory rate, and pain level) follow the same patterns. So, simply expanding the list of terms broadened the scope of the extraction tool. In order to take full advantage of the expanded scope, another set of 2,000 documents (1,000 training and 1,000 validation) was manually annotated using a broadened list of vital signs.

Additional analysis of vital sign terms and value co-occurrences uncovered that height, weight, pain level, and single values of either systolic or diastolic blood pressure never appeared in text without an associated term, or in very rare cases at least a unit of measure in a close proximity to the value. On the other hand, combined blood pressure (in "120/80" format), heart rate, respiratory rate, temperature, and oxygen saturation may appear as a stream of numbers without the associated terms in context.

The final system included modules to perform the following steps: detecting mentions of vital sign terms and units of measure; detecting numeric values; linking terms and values using patterns; and applying valid range filters on values of each vital sign type.

Results

The system was validated using 1,000 manually annotated documents, distinct from the training set. The tool performance is described in Table 1.

Table 1– Tool performance validated on 1,000 clinical notes

Vital	Recall	Precision
BP Diastolic	0.889	0.919
BP Systolic	0.858	0.927
Temperature	0.879	0.998
Heart rate	0.776	0.917
Pain level	0.454	0.993

Vital	Recall	Precision
Respiratory rate	0.766	0.956
O2 saturation	0.956	0.787
Height	0.800	0.800
Weight	0.824	0.999

Conclusion

Despite the large variability in clinical note content and formats, vital signs tend to appear in a relatively small set of patterns in the inpatient setting, and therefore rule-based extraction of values is both feasible and highly accurate.

References

[1] Cornia R, Patterson O V, Ginter T, Duvall SL. Rapid NLP Development with Leo. AMIA Annu Symp Proc. 2014.

Address for correspondence

olga.patterson@utah.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1036

Translating ICD-11 into French using lexical-based approach: a preliminary study

Tayeb Merabti[a], Julien Grosjean[a], Jean-Marie Rodrigues[b,c], Stefan Jacques Darmoni[a,b]

[a] *CISMeF & TIBS, LITIS EA 4108, Rouen University Hospital, Rouen, France*
[b] *INSERM, U1142, LIMICS, F-75006, Paris, France*
[c] *Department of Public Health, Saint Etienne University Hospital, France*

Abstract

To translate the 11th edition of the International Classification of Diseases (ICD-11) into French, we proposed a lexical approach using Natural Language Processing techniques. This method relies on the 56 biomedical terminologies and ontologies included in the Cross-lingual Health Multiple Terminologies and Ontologies Portal. From a sample of 336 ICD-11 terms, the algorithm translated 164 (49%) terms into at least one French term each.

Keywords:
Coding System, Mapping, Multilingualism, Semantic Interoperability, Terminology as Topic.

Introduction

The 11th edition of the International Classification of Diseases (ICD-11) is ongoing and its publication is not expected until 2017 [1]. Currently, ICD-11 is neither tranlated into French nor included in the Unified Medical Language System (UMLS). Thus, we have attempted to apply a lexical approach to translate each term from a sample of ICD-11 terms into at least one French term.

Materials and Methods

To translate ICD-11 terms, we proposed a lexical approach using Natural Language Processing (NLP) techniques. This method relies on 56 biomedical terminologies and ontologies (BMTO) included in the Health Multiple Terminologies and Ontologies Portal (HeTOP) [2]. Compared to UMLS, which includes 164,071 French terms, HeTOP includes 349,311 French terms[1] from BMTO.

This approach [3] helps to normalize all the English terms from the bilingual BMTO (English and French) included in the HeTOP. An algorithm was developed to find the terms most lexically similar to the target terms in the BMTO. When a correspondance was found, it was proposed as a possible translation of the English ICD-11 target term. The normalization process involved stripping genitive marks, transforming plural forms into singular, replacing punctuation, removing stop words, lower-casing each word, breaking a string into its constituent words, and sorting the words into their alphabetic order (see Figure 1). Examples of two ICD-11 term translations are listed in Table 1.

[1] Statistics calculated on preferred terms.

Table 1 - Examples of translation of two terms using lexical approach

ICD-11 term	English term (BMTO)	French term
Intracerebral hemorrhage	intracerebral hemorrhage (MeSH)	hémorragie cérébrale
Hypertensive encephalopathy	Hypertensive encephalopathy (MedDRA)	Encéphalopathie hypertensive

Figure 1 - Example of Normalization process for the ICD-11 term "Hereditary cerebral hemorrhage with amyloidosis, Dutch type"

Results & Discussion

According to our lexical approach, of the 336 ICD-11 terms selected, 164 (49%) terms were mapped from HeTOP and translated into at least one French term each. It is noteworthy that 194 (57%) terms were mapped to at least one English term. However, fifteen of these 164 translated ICD-11 terms (9%) were obtained exclusively from the translations performed by the CISMeF team [3].

Acknowledgement:

The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for editing assistance.

References

[1] http://www.who.int/classifications/icd/revision/en/ [accessed in November 2014].

[2] Grosjean J, Merabti T, Dahamna B, Kergouraly I, Thirion B, Soualmia L, Darmoni SJ. Health multi-terminology portal: a semantic added-value for patient safety. In: PSIP Workshop 2011; pp. 129-138.

[3] Merabti T, Soualmia LF, Grosjean J, Joubert M, Darmoni SJ. Aligning Biomedical Terminologies in French: Towards Semantic Interoperability in Medical Applications. In Book: Medical Informatics 2012; pp. 41-68.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1037

Text Mining and Data Modeling of Karyotypes to aid in Drug Repurposing Efforts

Zachary B. Abrams[a], Andrea L. Peabody[a], Nyla A. Heerema[b] and Philip R. O. Payne[a]

[a] *Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio, USA*
[b] *Department of Pathology, College of Medicine, The Ohio State University, Columbus, Ohio, USA*

Abstract

Karyotyping, or visually examining and recording chromosomal abnormalities, is commonly used to diagnose and treat disease. Karyotypes are written in the International System for Human Cytogenetic Nomenclature (ISCN), a computationally non-readable language that precludes full analysis of these genomic data. In response, we developed a cytogenetic platform that transfers the ISCN karyotypes to a machine-readable model available for computational analysis. Here we use cytogenetic data from the National Cancer Institute (NCI)-curated Mitelman database1 to create a structured karyotype language. Then, drug-gene-disease triplets are generated via a computational pipeline connecting public drug-gene interaction data sources to identify potential drug repurposing opportunities.

Keywords:

Cytogenetics; Karyotype; Text mining; Drug repurposing

Introduction

Cytogenetic data in the form of karyotypes are commonly used in the diagnosis and treatment of many forms of cancer. Karyotype data are expressed in a text-based form that is not machine-readable. This limits the utility of these data for secondary use and research purposes. Utilizing the International System for Human Cytogenetic Nomenclature (ISCN), we developed a parsing and mapping system that allows karyotype data to be represented and analyzed in a computationally tractable manner. A Loss-Gain-Fusion model (LGF) was created that allowed us to represent each karyotype as a binary vector. Each cytogenetic region is represented three times (loss, gain, and fusion) in the model. We utilized the publicly available Mitelman database as a test-bed for analyses, focusing on problems related to drug repurposing.

Materials and Methods

Utilizing our computational model and the Mitelman database, we were able to successfully parse 98% of its karyotypes; of those parsed, 89.4% could be mapped into our binary Loss-Gain-Fusion model. We then classified karyotypes based on their disease labels and filtered out all diseases with less than 50 patients. We then selected genetic aberrations present in 20% or more of the population in which the cytogenetic event led to increased gene expression. Subsequently, we identified all genes in the affected region and found drugs that inhibited the function of the overexpressed gene using publicly available drug data in The Drug Gene Interaction Database (DGIdb). We performed a literature search on these results in PubMed selecting diseases and drugs that did not co-occur and where the disease and the gene had co-occurred in at least one PubMed abstract.

Results

We discovered 68,543 triplets containing (1) a disease, (2) an overexpressed gene, and (3) a drug that suppressed that specific gene. From this list, we discovered a total of 69 cancer disease-drug pairs that were not cited as co-occurring in the literature. Given this filtering process where the drug and gene are related, the drug suppressed the gene and the gene was implicated in the disease; it logically follows that the drug should be helpful in treating the disease.

Discussion

Our computational approach serves as a basis for new directions in drug repurposing, leveraging existing and commonly available bio-molecular phenotypic data. In order to validate our results, future laboratory-based testing will be conducted on a sub-set of our findings. The ability to link publicly available data sources is a central component of this work and emphasizes the importance of utilizing such data in conjunction with clinically-generated data sets so as to support in-silico hypothesis generation.

Conclusion

Utilizing publicly available data sets, we generated a list of 68,543 drug-gene-disease triplets, each triplet containing a gene that is up-regulated, a drug that works to suppress or inhibit that up-regulated function, and a disease where the up-regulated gene is an implicated disease agent. This information may play a significant role towards drug repurposing efforts in the 69 diseases. We anticipate that this information will prove useful to researchers in the domain of pharmaceutical drug repurposing and in the treatment of these conditions.

References

[1] Mitelman F, Johansson B, Mertens F. Mitelman Database of chromosome aberrations and gene fusions in cancer [Internet]. 2014.. Available from: http://cgap.nci.nih.gov/Chromosomes/Mitelman

Address for correspondence

350 Lincoln Tower 1800 Cannon Dr. The Ohio State University Columbus, OH 43210 United States

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1038

Rule-based Cervical Spine Defect Classification Using Medical Narratives

Yihan Deng[a], Mathias Jacob Groll[b], Kerstin Denecke[a]

[a] *Innovation Center of Computer Assisted Surgery, University of Leipzig, Germany*
[b] *Neurosurgical Department, University Hospital Leipzig, Germany*

Abstract

Classifying the defects occurring at the cervical spine provides the basis for surgical treatment planning and therapy recommendation. This process requires evidence from patient records. Further, the degree of a defect needs to be encoded in a standardized form to facilitate data exchange and multimodal interoperability. In this paper, a concept for automatic defect classification based on information extracted from textual data of patient records is presented. In a retrospective study, the classifier is applied to clinical documents and the classification results are evaluated.

Keywords:

Information Extraction; Cervical Vertebrae; Knowledge Bases.

Introduction

The degenerative changes of the cervical spine are reflected in multiple aspects of observation such as area of the defect, position of the defect and additional pathology. However, two challenges should be solved to enable the automatic defect classification in surgical context. Firstly, although the extraction of clinical evidence or disease characteristics from texts has been studied thoroughly using both rule-based or machine learning-based methods [1], the development of domain-specific extractor and classifier with high accuracy is still necessary for cervical spine diseases. Secondly, most of the available schemas for cervical spine defect classification focus on the pathological changes in the spinal canal or base on the cross-sectional area. A surgical classification schema must be implemented to provide surgeons with most essential defect situations.

Methods

In this paper, we employed a novel surgical oriented classification schema for the grading of spinal canal defects. It shows more defect information and categorizes additional pathological information for surgical practice use. It covers the amount of the affected cervical segments (Mono, Bi, Tri+, Skip Lesion), position of the defect (Medial (m), Lateral (l), Medial and Lateral (m&l)) and additional pathologies (Thickening of ligaments (1), disorders (listhesis, kyphosis, etc.) (2)). E.g., “bi-m2” refers to a bi-segmental stenosis with medial compression of the spinal cord with additional disorder.

A rule-based classification system based on extracted features are developed. As can be seen in figure 1, the information entities in medical narratives are extracted through concept mapper and regular expressions, while the classification schema is transformed into knowledge rules. As next, the rule based classifier determines the concrete defect classes for the patient record. Three hundred and eleven defect terms are categorized and summarized in the terminology manually. The relevant terms in the terminology list were directly linked to existing German terminologies such as Radlex, ICD 10 and MeSH 2010 as expansion. The defect relevant concepts, diagnoses and anatomical concepts are obtained through fuzzy matching and exact matching. The date, negation, measurement and dose in unit are also extracted. Fifty four facts and seven rules have been defined in our knowledge base using Prolog[1].

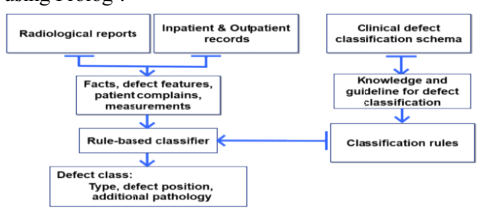


Figure 1- System flow chart of defect classification based on patient records

Results

One hundred samples of patient records in German were annotated with their defect features and classified defect codes as benchmarks by one physician. The extraction of the three defect features has reached high F1 measures ranging from 91% to 99%. The defect classification has also achieved generally high accuracy except for the detection of the position of “medial” (53%) due to an insufficient rule of facts in the knowledge base.

Conclusions

As a pilot study, preliminary experiments have been conducted with the support of surgeons. The usefulness of the automatic classification was confirmed by clinical experts. An additional clinical study will be organized for collecting more rule defintions from empirical experience. The rules will also be evaluted based on more patient record samples.

References

[1] Coden A, Savova G, Sominsky I, Tanenblatt M, Masanz J, Schuler K, Cooper J,Guan W, de Groen PC. Automatically extracting cancer disease characteristics frompathology reports into a Disease Knowledge Representation Model. J Biomed Inform 2009: Oct 42(5):937-49.

Address for correspondence

yihan.deng@iccas.de

[1] http://www.swi-prolog.org/

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1039

Comparing Drug-Disease Associations in Clinical Practice Guideline Recommendations and Drug Product Label Indications

Tiffany I. Leung, MD, MPH[a], Michel Dumontier, PhD[b]

[a]*Division of General Medical Disciplines, Stanford University, Stanford, CA*
[b]*Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA*

Abstract

Clinical practice guidelines (CPGs) and structured product labels (SPLs) are both intended to promote evidence-based medical practices and guide clinicians' prescribing decisions. However, it is unclear how well CPG recommendations about pharmacologic therapies for certain diseases match SPL indications for recommended drugs. In this study, we use publicly available data and text mining methods to examine drug-disease associations in CPG recommendations and SPL treatment indications for 15 common chronic conditions. Preliminary results suggest that there is a mismatch between guideline-recommended pharmacologic therapies and SPL indications. Conflicting or inconsistent recommendations and indications may complicate clinical decision making and implementation or measurement of best practices.

Keywords:

Clinical Practice Guidelines; Drug Labeling; Drug Therapy; Information Storage and Retrieval; Chronic Disease.

Introduction

Clinical practice guidelines (CPGs) make recommendations on pharmacologic treatments for clinical conditions, and drugs' structured product labels (SPLs), provided by the U.S. Food and Drug Administration, summarize approved treatment indications. Both resources are intended to promote evidence-based medical practices and guide clinicians' prescribing decisions. Ideally, CPG and SPL content should support similar prescribing practices, reflecting a similar evidence base, however, no study has been done to evaluate how well these resources match. It is also labor-intensive to perform a manual review of the numerous existing CPGs and SPLs. In this study, we use text mining methods to examine how well CPG recommendations about pharmacologic treatment options for 15 common chronic conditions match with SPL treatment indications. We hypothesize that CPG recommendations and SPL treatment indications should be high.

Methods

We focused on 15 common chronic conditions among Medicare beneficiaries, using 448 ICD-9 codes to identify 494 relevant guideline summaries from the National Guideline Clearinghouse to build the CPG text corpus. To construct drug-disease associations from the CPG text corpus, we applied a text-mining algorithm that performs named entity recognition, similar to that used in a previous study to extract chronic disease mentions (1); however, we searched for drug names in the CPG text, using a drug list of over 12,000 drug names and synonyms from DrugBank, which we then enhanced with disease and drug vocabularies from Bioportal at Stanford's National Center for Biomedical Ontology. Then, we built drug-disease associations from SPLs' Indications section using a computer-readable side effect resource (SIDER). We measured the frequencies of matches and mismatches of drug-disease associations in CPGs and SPLs.

Results

We identified 1,998 unique drug-disease associations from CPGs; in SPLs, 533 unique drug-disease associations were identified. Of these, only 240 (45%) of these drug-disease associations matched between CPGs and SPLs. This means that CPGs mention 1,758 (88.0%) drug-disease associations that are not also mentioned in SPLs; conversely, SPLs mention 293 (55.0%) drug-disease associations that are not also mentioned in CPGs.

Conclusions

Our results suggest that there are mismatches between guideline-recommended pharmacologic therapies and SPL indications, suggesting potentially conflicting prescribing guidance from CPGs and SPLs. This can pose practical challenges for clinicians, patients and caregivers, and practice administrators. Further investigation of the identified mismatches may guide harmonization of clinical knowledge sources. In future work, improvement and evaluation of our text mining methods may address limitations of the present approach and extract relationship information between drug and disease.

Address for correspondence

Tiffany I. Leung, MD, MPH, 211 Quarry Road, Suite #102, Palo Alto, CA 94305. E-mail: tileung@stanford.edu

References

[1] Leung TI, Jalal H, Zulman DM, Dumontier M, Owens DK, Musen M, Goldstein MK. Automating Identification of Multiple Chronic Conditions in Clinical Practice Guideline recommendations. *AMIA Jt Summits Transl Sci Proc*. 2015; Accepted for publication.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1040

Automated Classification of Pathology Reports

Michel Oleynik[a], Marcelo Finger[a], Diogo F. C. Patrão[b]

[a] *Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil*
[b] *International Center for Research, A. C. Camargo Cancer Center, São Paulo, Brazil*

Abstract

This work develops an automated classifier of pathology reports which infers the topography and the morphology classes of a tumor using codes from the International Classification of Diseases for Oncology (ICD-O). Data from 94,980 patients of the A.C. Camargo Cancer Center was used for training and validation of Naive Bayes classifiers, evaluated by the F_1-score. Measures greater than 74% in the topographic group and 61% in the morphologic group are reported. Our work provides a successful baseline for future research for the classification of medical documents written in Portuguese and in other domains.

Keywords:

Natural Language Processing; Medical Records; Data Mining.

Introduction

Pathology reports are an important data source for diagnosis in the cancer domain because reports provide clinical evidence on the topography of the tumor, its histological type, and morphology. The information may be encoded with the International Classification of Diseases for Oncology (ICD-O). A recent work [1] evaluated the use of Naive Bayes classifiers and Support Vector Machines, and we applied a modified version of these techniques on pathology reports written in Portuguese and assessed its efficiency.

Methods

We used a collection of pathology reports written in Portuguese from the A.C. Camargo Cancer Center in São Paulo, Brazil. The dataset comprises 94,980 patients from the years 1196 to 2010.

We applied a multinomial Naive Bayes classifier [2], seen in Figure 1, with add-alpha smoothing to the resulting data (we chose $\alpha = 0.5$ since our belief in uniformity is weak).

$$c_{map} = \underset{c \in C}{\operatorname{argmax}} [\log \hat{P}(c) + \sum_{1 \le k \le n_d} \log \hat{P}(t_k \mid c)]$$

Figure 1 – Naïve Bayes Classifier

Results

While one work [1] reported 71.5% F_1-measure on the classification of 26 topographic groups, we presented a higher F_1-measure (75.1%) on a task of 16 groups. On the morphology axis of the ICD-O classification, we achieved a lower F_1-measure (61.3% *versus* 85.4%) on a wider classification task (49 *versus* 18 groups). The results were micro-averaged with 10-fold cross-validation.

Table 1 – Micro-averaged efficiency (%) of Naive Bayes classifiers trained on increasing levels of detail of the ICD-O

	Topography			**Morphology**		
	P	**R**	**F_1**	**P**	**R**	**F_1**
Group	78.928	73.451	75.304	68.813	60.172	62.256
Category	73.180	69.284	69.284		N/A	
Concept	43.239	40.820	39.941	49.151	39.705	40.673

When analyzing the classifier efficiency on the topographic group, we achieved 98.3% and 97.3% precision on the groups *C50: Breast* and *C60-C63: Male genital organs*, respectively the most common cancer for women and men. We also analyzed the resulting confusion matrix and observed that the most common cause of error is the incorrect classification of *C50: Breast* as *C51-C58: Female genital organs*. This is probably due to the fact that the diagnosis of breast cancer is usually accompanied by a screening test for cervical cancer.

Conclusion

Our work has immediate implications to knowledge discovery and statistical analysis in the medical domain, as it eases the process of obtaining structured information over textual data. It also accelerates the work of physicians when classifying patient data.

Acknowledgments

This research was funded by the Brazilian National Research Council - CNPq (project number 133862/2010-0) and is approved by the committee on ethics on research of the A.C. Camargo Cancer Center (registered under number 1418/10).

References

[1] Jouhet V, Defossez G, Burgun A, Le Beux P, P. Levillain, et al. Automated classification of free-text pathology reports for registration of incident cases of cancer. Methods of Information in Medicine. 2012;51(3):242.

[2] Pearl J. Bayesian Networks: a model of self-activated: memory for evidential reasoning. Computer Science Department, University of California; 1985.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1041

Methods for Sonic Representation of ST Depression During Exercise

Minodora Andor[a], Anca Tudor[a], Sorin Paralescu[a], George I Mihalas[a,b]

[a] *Victor Babes University of Medicine and Pharmacy, Timisoara, Romania*
[b] *Vasile Goldis Western University, Arad, Romania*

Abstract

Sonic display of ST depression during exercise helps both the patient and the investigator to better identify the transition from normal values to "attention" region and reaching the "alert" threshold. Two types of sonic display were tested, based on combinations of saccadic sounds and different pitch.

Keywords:

Sonification; ST depression; HR. alert threshold.

Introduction

This work is part of a larger project on "Adding Sound to Medical Data", to develop a toolkit for sonification of biological signals [1]. This study describes the module (ST) of the ASET package (Adding Sound to Exercise Test). Modern equipment displays the parameter evolution during exercise, underlying critical values of blood pressure (BP), heart rate (HR) and ST segment depression, each with well defined values as endpoints. The information is only visually displayed, the patient is kept passive. Our purpose was to improve the parameter perception and involve the patient.

Methods

Signals and Parameters

Signals. We used both our recordings and signals from Physiobank [2], from two leads (AVF or III and V5 or V6). The ST segment level was computed at the J point [3].

Thresholds. We used a simplified risk representation, preserving two major levels of warning, attention (close to risk threshold) and alert (endpoint), which are visually represented by a light red background of that parameter. The alert threshold STb = -1 mV in V5/V6 or STb = -1.5 mV in AVF/III , and the attention threshold was STa = 0.8 STb.

Sonification procedures

From the large selection of possible sonic representations, we have selected the following criteria: *duration* - equal to the RR interval for each beat, *pitch* - the reference frequency (f_0) was 523.25 Hz (C5 on musical scale), corresponding to ST_i < STb. For the three sounds displays, the next coefficients were ka=6/5; kb= 4/3, and *saccadic sounds.* We tried both continuous and saccadic sounds; a "saccade" was defined as a short sound (0.1 RR) followed by a pause of 0.02 RR. It can represent warning levels: 1 saccade for attention and three saccades for alert.Two types of sonic display were tried: saccadic sounds only ("1S") – no pitch variation (all C5), and saccadic display ("3S") of three sounds (C5/D5/E5): no saccade on f_0 - for normal range of HR, 1 saccade on $f_0 \times$ ka for attention level and 3 saccades on $f_0 \times$ kb for alert.

Preference selection

We conducted a survey with a group of 12 subjects. The survey had two phases: *supervised learnin*g - the subjects were exposed to the sounds and explained the two representations (named 1S and 3S), and *testing phase* - the subjects had to listen and score each record with marks between 10 and 0 (10 for perfect distinction). Each record was trimmed from the original 7-30 min. to 30–90 sec., preserving the transitions. The ASET package consists of two main scripts (in MATLAB 2011b) and several functions: ECG signal import into MATLAB, detection of RR intervals, detection of J point, ST value, and minimal ST level of the two leads. Then each signal was transformed for the two sonic representations (1S or 3S), using a function for saccades .

Results and Discussion

The sonic files used for testing with their graphical display and the corresponding sounds can be found on our website [4] in the section ASET. For stereo systems, HR sounds are sent to the right speaker while extra systoles are sent to the left.Our results (from 12 subjects, each for 5 signals) show that all sonic representations used in this study have very good discriminant power but different preference: the 1S display was not preferred when the 3S display was also available, but it can be kept in simple devices. From the comments made by the subjects we discovered preference to omit any sound before reaching the attention threshold and to introduce a dissonant note to mark the alert.

Conclusion

Our study showed a clear preference for simple displays (with no details), focused on revealing the threshold crossing. The potential impact of our study goes beyond the monitoring of cardiac performance during the exercise tests, to wearable devices with simple sonic warning system.

References

[1] Mihalas GI, Paralescu S, Andor M et al. Tempolenses with Variable Magnification in Sonification of Cardiac Signals. *Stud Health Technol Inform*. 2013, **186**:78-82.
[2] Goldberger AL et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation,*2000, **101**(23):e215-e220.
[3] Brownfield J; Herbert M. EKG Criteria for Fibrinolysis: What's Up with the J Point? *The western journal of emergency medicine* 2008, **9** (1): 40–2.
[4] www.medinfo.umft.ro/dim/sonification/index.htm.

Address for correspondence

George Mihalas, Vasile Goldis Western University, Arad, Romania, mihalas@uvvg.ro.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1042

PIACS: A System for the Automatic Detection, Categorization and Comparison of Scratch-Related Skin Lesions in Dermatology

Philipp Bruland[a], Waldemar Hänse[a], Fiona Schedel[b], Sonja Ständer[b], Fleur Fritz[a]

[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Dermatology, University of Münster, Germany*

Abstract and Objective

In the treatment of chronic pruritus-related, scratch-induced skin lesions the categorization, counting and temporal comparison are common methodologies. The observation requires a good memory and expertise in this field to gain comparable findings for this time-consuming process. Digital image processing aims at supporting such manual detections. The objective is to develop a software tool for automatic image detection and comparison. The new photographic setting implies the usage of markers to derive the brightness and size of lesions. MATLAB has been used for the software development. The newly defined setting allows taking standardized images of pruritus-associated cutaneous lesions for detection and comparison. The tool named PIACS (Prurigo Image Analyzing and Comparing System) allows automatically detecting, categorizing and comparing lesions based on digital images.

Keywords:

Image processing; digital image detection; prurigo nodularis.

Introduction

The diagnostic and treatment of chronic pruritus becomes an increasingly important medical field in the dermatology. Besides common skin diseases, scratch-related skin lesions frequently occur in pruritus patients [1]. These lesions show a broad variability in redness, circularity as well as size and the also may serve as a monitor marker in the course of the therapy; however, they can only tediously be counted manually. Several hundred (mean, 180±146) of lesions are not uncommon. The manual process of classifying and counting is very time-consuming and not applicable in daily routine or clinical trials where a standardized procedure would be expedient [2].

Images are taken in a photographic laboratory but the photographic equipment has to be modified quite often. Flashes and lenses are changed, resulting in inconsistent brightness of the images. Automatic image detection is mainly applied for nevi and melanoma assessment for which images are taken [3]. The objective of this work is to develop a software tool to detect, classify and compare scratch-related skin lesions.

Methods

MATLAB was chosen for the development of since it offers the most comprehensive collection of image processing methodologies. For the three different lesions predefined parameters were applied for each category. Round gray markers were used to derive the real size in the image. The results of PIACS were evaluated by regarding correctness and accuracy of the detection within 25 randomly selected images.

Results

For the automatic detection of scratch lesions the following steps are performed: To extrapolate the skin and the marker in the image, it is separated into background, skin and the round marker. After that the marker is detected and the size derived for one pixel calculated in millimeter. The grey marker is used for color adjustment. Therefore, a reference image is prepared through the creation of red, green and blue histograms from 30 images. For the analyzed image the histogram adjustment is applied for each color channel. The detection of skin lesions is achieved by the subtraction of color channels and thresholding. The expanse and category is determined by the application of predefined parameters for each category. For the comparison two images are analyzed and the differences calculated.

On the 25 selected images in total 692 lesions were assigned. In total, 493 are correctly detected, 175 are categorized wrong and 24 are not detected. This results in an overall sensitivity of 95.7% and an accuracy of 75.3%.

Discussion / Conclusion

The examination process of pruritus-related scratch lesions may benefit from such an approach and is urgently needed for the application in diagnostic and treatment as well as clinical trials to allow proper and objective assessment of the course of the symptom. The accuracy might be improved through the introduction of artificial neural networks.

References

[1] Ständer S, et al. Clinical classification of itch: a position paper of the International Forum for the Study of Itch. Acta Derm Venereol. 2007;87(4):291-4.

[2] Allen AM. Clinical trials in dermatology, part 3: measuring responses to treatment. Int J Dermatol. 1980 Jan-Feb;19(1):1-6.

[3] Blum A, et al. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. Br J Dermatol. 2004 Nov;151(5):1029-38.

Corresponding author: Philipp.Bruland [at] uni-muenster.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1043

The MFER structure for coding medical signals in real time

Yonghee Lee[a], Jaehyuk Kim[a], Soonseok Kim[a], Kangwoo Lee[a], Dongho Kim[a]

[a] *Department of Computer Engineering, Halla University, Wonju, Korea*

Abstract

We present the structure of the MFER coder to store the medical signals in real time. The MFER is the rules to describe the file with tags the medical signal such as ECG, EEG, etc. However, because the MFER has the simple structure, and is stored as a file unit, it is difficult to process the signal with MFER rules in real-time. To do this, By storing the signal into small unit of time, it is possible to handle as if the signal is stored continuously. The structure of the presented MFER coder is verified in the results.

Keywords:
medical waveform, MDER, healthcare, MFER

Introduction

The MFER (Medical Waveform Format Encoding Rules) is a set of rules to store the medical waveforms such as ECG, EMG, EEG, etc., its rules basically consist of the tag (T), the length (L), and length of data value (V) based on MDER. And, the value (V) represents data indicated by the tag (T). To describe a signal, the MFER consists of the sampling attribute, frame attribute, and supplementary information. The file data is fundamentally composed of a header and waveform data of the signal, and the header and waveform data are encoded after dividing them into a tag, length, and value. However, even though the MFER has the simple structure so that we are able to deal with them, easily, the structure does not support real-time process. It is very important to store and monitor the waveform of a patient in real time from a remote location over the network. In this paper, we propose the coding structure to monitor in real-time based on the basic MFER.

Materials and Methods

The coding structure

Figure 1 shows the structure of the proposed MFER coder. First, We send the header block containing the waveform information before transmitting the continuous wave from the device to the server. This process will automatically transfer from the medical device, or enter manually in server system. Then, the medical device send a continuous waveform, and the server is attached to a block of the header information by dividing the waveform signal into block units. We are able to deal with, such as if subsequently store and monitor the waveform information when the monitoring result from the server side. Figure 2 describes how to design the corresponding tag. This interface could make quite good use when tags are called or discriminated in a linked list later.

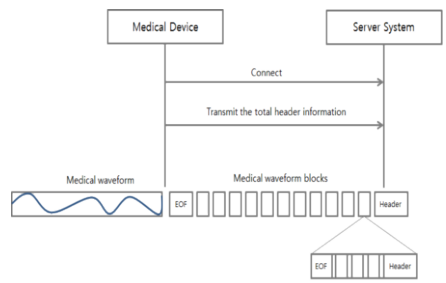


Figure 1 - The coding structure

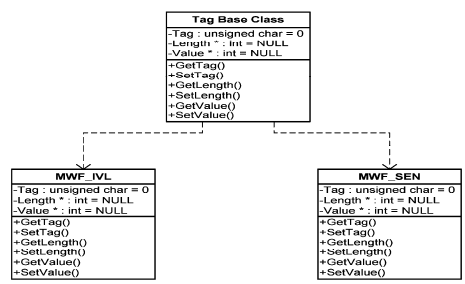


Figure 2 – Design of Tags

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2015-H8601-15-1009) supervised by the IITP (Institute for Information & Communications Technology Promotion)

References

[1] Kimura, E., Development MFER (Medical waveform Format En-coding Rules) Encoder, AMIA Annual Symposium Proceedings, 2006.

[2] M Hirai, K Kawamoto. MFER-a Japanes Approach for Medical Wave Form Encoding Rules for Viewer Design, OpenECG Work-shop, Berlin, Germany, pp. 35-37, 2004.

Address for correspondence

yhlee@halla.ac.kr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1044

Adding Sound to ECG

George I Mihalas[a,b], Sorin Paralescu[a], Anca Tudor[a], Minodora Andor[a]

[a] *Victor Babes University of Medicine and Pharmacy, Timisoara, Romania*
[b] *Vasile Goldis Western University, Arad, Romania*

Abstract

This poster presents preliminary results of a project aiming to develop tools for adding sound associated to medical data for potential medical applications. Sonification procedures, the methodology used for testing various sonic representations of ECG, and the results are presented.

Keywords:

Sonification; Data representation; ECG; Discriminant power.

Introduction

The use of sonification for representation of medical data, especially biosignals is not a new idea [1], but most previous work has limited approaches. A systematic study has yet to be completed, due to the diversity of methods to "sonify" a set of data. From this premise, we initiated a project: “Adding Sound to Medical Data”, with four steps [2]:

- preparing the sonification tools and methodology;
- testing the discriminant power of various sets of parameters;
- developing the procedure for clinical application;
- development of sonification modules for medical equipment.

This paper refers only to the first two phases.

Methods

The first stage aimed to reduce the degree of arbitration in mapping the physical parameters of the original signal (data).

Sonification procedures

In our approach [3], the relation between the sound pitch (f_i) and normalized data [0,1] of signal amplitude y_i was:

$$f_i = f_0 \times 2 \text{ ^ } (y_i),\ f_0 = 523.25 \text{ Hz (note C5)} \quad (1)$$

Three major sonification levels could be defined:

- acoustic level, with a continuous frequency spectrum (f);
- sonic level (S) – with discrete spectrum, from musical scale;
- musical level (M) – multichannel, introducing rhythm and harmony. Level M will not be referred to in this paper.

We split the acoustic level into two (sub)levels:

- *continuous* representation, called (sub)level (A): for two neighbor points (t_i, f_i) and (t_{i+1}, f_{i+1}), the frequency will vary continuously from f_i to f_{i+1};
- *quasicontinuous* (sub)level (Q) representation: only the frequency f_i will be produced for the interval dt = (t_i, t_{i+1}), followed by f_{i+1} for the next interval dt , etc.

We used signals from PhysioBank [3] (sampling rate 250 Hz) from both healthy subjects and from patients with sleep apnea, arrhythmia and congestive heart failure. Additionally, we built a tool similar to a lens - tempolens, (TL) able to dilate or compress temporal sonic display, with fixed (f) or variable (v) magnification [2].

Results and Discussions

Computer programs

Our package was built on MATLAB R2011b platform, for sonifying a signal in levels A, Q and S and applying tempolenses. An example of the screen produced for an ECG signal of 10 seconds and the sounds can be displayed by accessing reference [4].

Discriminant power

We tested the discriminant power (the capacity of listeners to distinguish details and recognize signals). Our results showed:

- low preference for A mode (it sounds like a whistle), but had a high discrimination in sleep apnea (obstructive episodes);
- tempolenses with variable magnification did not bring the expected increase in resolution of the QRS complex;
- durations of 0.2 seconds or less in Q mode sounded like A.

Conclusions

We built and tested a methodology for adding sound to medical data to improve data analysis in various cases: HR variability recognition, differential diagnosis, exercise tests, help people with visual impairment, etc.

References

[1] Ballora M et al. Heart Rate Sonification: a New Approach to Medical Diagnosis. *Leonardo,* 2004; **37**(1):41-46.
[2] Mihalas GI, Paralescu S, Andor M et al. Tempolenses. *Stud Health Technol Inform.* 2013, **186**:78-82.
[3] Goldberger AL et al PhysioBank, PhysioToolkit and Physio Net Components. *Circulation* 2000; **101**(23):e215-e220.
[4] http://www.umft.ro/dim/sonification.

Address for correspondence:

Prof. George I Mihalas, Vasile Goldis Western Univ, Bd. Revolutiei 94, Arad, Romania, mihalas@uvvg.ro.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1045

An Optimized Superpixel Clustering Approach for High-Resolution Chest CT Image Segmentation

Rafaelo Pinheiro da Rosa[a], Marcos Cordeiro d'Ornellas[a]

[a] *Laboratório de Computação Aplicada, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*

Abstract

Lung segmentation is a fundamental step in many image analysis applications for lung diseases and abnormalities in thoracic computed tomography (CT). However, due to the large variations in pathology that may be present in thoracic CT images, it is difficult to extract the lung regions accurately. A major insight to deal with this problem is the existence of new approaches to cope with quality and performance. This poster presents an optimized superpixel clustering approach for high-resolution chest CT segmentation. The proposed algorithm is compared against some super-pixel algorithms while a performance evaluation is carried out in terms of boundary recall and under-segmentation error metrics. The over-segmentation results on a CT Emphysema Database demonstrate that our approach shows better performance than other three state-of-the-art superpixel methods.

Keywords:

Image Segmentation; Radiography Image Interpretation; Tomography.

Introduction

Medical imaging is now used to provide interventional treatments and diagnostic screening. It allows for substantial information about abnormal adjacent tissues, helping physicians to identify the appropriate treatment. Recently, superpixels have converted into an essential approach for many imaging applications including segmentation [1]. This poster presents an optimized superpixel clustering approach for high-resolution chest CT segmentation. The proposed algorithm is compared against some superpixel algorithms while a performance evaluation is carried out in terms of boundary recall and under-segmentation error metrics.

Methodology

The algorithm presented in this poster is an optimized version of the original SLIC algorithm [2], with some differences. A performance evaluation of the ESLIC algorithm was exhibited by comparing its accuracy against SLIC, NC00, QS08, and TP09 algorithm. The ESLIC algorithm was implemented in Java. While only comparisons with original algorithms and extensions have been made, computer benchmarks were performed in C using a commodity computer.

Results

Figure 1 in the left shows the results obtained with under-segmentation error metric. It can be concluded from this plot that ESLIC exceeds the other algorithms after 900 superpixels. In the right, results for boundary recall are displayed. In this plot, ESLIC exceeds QS08 and TP09. It is also higher than NC00 when 50 superpixels are considered. Figure 2 shows a visual comparison, produced by each one of the algorithms considered in this work.

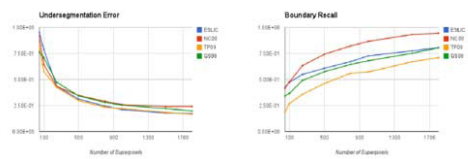


Figure 1 – Plot of USE (left) and BR(right) with respect to number of superpixels.

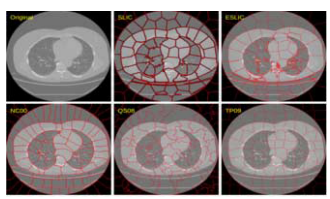


Figure 2 – Segmentation results on CT sample for SLIC12, ESLIC, NC00, QS08, and TP09.

Conclusion and Further Work

In this poster, NC00, QS08, TP09, and SLIC12 algorithms were compared against ESLIC. Results have shown that the proposed algorithm is superior, with respect to under-segmentation error, to the other algorithms. When considering boundary recall, ESLIC performs better than QS09 and TP09 but is outperformed by NC00.

References

[1] Bergh M, Boix X, et al. SEEDS: Superpixels Extracted via Energy-driven Sampling. European Conference on Computer Vision, pages 13–26, 2012.

[2] Achanta R, Shaji, A, et al. SLIC Superpixels. Technical report, École Polytechnique Fédérale de Lausanne, 2010.

Address for correspondence

Rafaelo Pinheiro da Rosa, Email: rafaelo.pinheiro1988@gmail.com.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1046

Automated Image Retrieval of Chest CT Images Based on Local Grey Scale Invariant Features

Marcelo Arrais Porto[a], Marcos Cordeiro d'Ornellas[b]

[a] *Animati Computação Aplicada, Santa Maria-RS, Brasil*
[b] *Universidade Federal de Santa Maria, Universidade Federal de Santa Maria, Santa Maria-RS, Brasil*

Abstract

Textual-based tools are regularly employed to retrieve medical images for reading and interpretation using current retrieval Picture Archiving and Communication Systems (PACS) but pose some drawbacks. All-purpose content-based image retrieval (CBIR) systems are limited when dealing with medical images and do not fit well into PACS workflow and clinical practice. This paper presents an automated image retrieval approach for chest CT images based local grey scale invariant features from a local database. Performance was measured in terms of precision and recall, average retrieval precision (ARP), and average retrieval rate (ARR). Preliminary results have shown the effectiveness of the proposed approach. The prototype is also a useful tool for radiology research and education, providing valuable information to the medical and broader healthcare community.

Keywords:

PACS; Image Retrieval; Chest CT Images; Local Features.

Introduction

Medical imaging has become essential for modern medicine. However, image interpretation and understanding poses a significant and continuous burden on physicians who need to evaluate large amounts of data. In order to tackle these challenges, a generic integration of computer-aided diagnosis (CAD) and content-based image retrieval (CBIR) is proposed. CBIR usually employs a set of low-level feature descriptors to represent a medical image, from which a group of similarity functions are used to drive different sorts of queries. This helps physicians to find similar cases from a variety of archives, thus providing support with medical image interpretation and decision-making.

This paper presents an automatic approach to image retrieval of chest CT images based on local grey scale invariant features such as scale-invariant feature transform and multi-scale oriented patches.

Methods

Feature vectors, obtained by low level features, are fundamental to assess similarity measurements in CBIR searching process. Feature extraction methods must be stable and robust to enable image recognition and retrieval. Local invariant methods have been employed to tackle these problems including SIFT and MOPS[1]. We adopt MOPS algorithm in this paper for feature extraction and matching of chest CT images. MOPS is a relatively lightweight scale invariant feature detector compared to SIFT, and has advantage of faster detection speed.

Results

A MOPS implementation was tested using three different public domain training datasets (CTED[2], TCGA-LUSC, and NSCLC-Radiomics) to validate its efficiency. The algorithm ranks the images in descending order of the maximum region similarity. The top ranked images are taken as the candidate retrieval results.

The retrieval performance of the proposed method is measured in terms of recall, precision[*], average retrieval rate (ARP) and average retrieval precision (ARR). Figures 1 illustrates the retrieval performance comparison of the proposed approach for all datasets in terms of ARP and ARR.

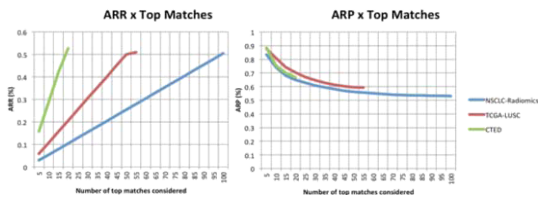


Figure 1 – ARPxTop (left) and ARRxTop (right) matches for all datasets.

Conclusion and Further Work

The goal of this paper was to provide an automated chest CT image retrieval approach based on local grey scale invariant features. To access the correctness of query results, an evaluation procedure was carried out in terms of PR plots, average retrieval precision and average retrieval rate. The proposed approach is still not optimal and performance considerations must be taken into account in the near future.

References

[1] Brown M, Szeliski R, and Winder S. Multi-Image Matching Using Multi-Scale Oriented Patches. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (1)510– 517, 2005.

[2] Sørensen L, Shaker SB, and de Bruijne M. Quantitative Analysis of Pulmonary Emphysema using Local Binary Patterns, IEEE Transactions on Medical Imaging 29(2): 559-569, 2010.

Address for correspondence

Marcelo Arrais Porto (marcelo@animati.com.br)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1047

Texture Analysis of Recurrence Plots Based on Wavelets and PSO for Laryngeal Pathologies Detection

Taciana A. Souza [a], Vinícius J. D.Vieira[b], Suzete E. N. Correia[a], Silvana L. N. C. Costa[a], Washington. C. de A. Costa[a], Micael A. Souza[a]

[a]Academic Unity of Industry, Electrical Engineering, IFPB, João Pessoa, Brazil
[b]Electrical Engineering Post-graduation Program, UFCG, Campina Grande, Brazil

Abstract

This paper deals with the discrimination between healthy and pathological speech signals using recurrence plots and wavelet transform with texture features. Approximation and detail coefficients are obtained from the recurrence plots using Haar wavelet transform, considering one decomposition level. The considered laryngeal pathologies are: paralysis, Reinke's edema and nodules. Accuracy rates above 86% were obtained by means of the employed method.

Keywords:

Wavelet Transform; Haralick Features; Recurrence Plots; Laryngeal Pathologies.

Introduction

Laryngeal pathologies might affect the speech production process. Recurrence plots have been used in the discrimination of healthy and pathological speech signals recently [1]. These structures show small scale patterns (which can be seen as texture features) and large scale patterns (which can be seen as typology features) [2].

This research employs Haralicktexture analysis features [3] on coefficients obtained from sub-images of the wavelet transform [4] applied in recurrence plots of speech signals. The discriminative power among pathologies in the larynx is investigated through the inspection of the visual information in the recurrence plots. Three pathologies are considered for analysis: paralysis, Reinke's edema and nodules.

Methods

Approximation and detail coefficients are obtained from the recurrence plots using Haar wavelet transform, considering one decomposition level. The texture analysis is carried out with Haralick features on these coefficients. The speech signals are classified by multilayer perceptron neural network (MLP) with cross validation. The most significative features are selected by particle swarm optimization (PSO) [5]. This classification step is carried out for each coefficient matrix (approximation, horizontal, vertical and diagonal details).

The speech signals used in this research belong to a database recorded by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [6]. The use of 53 signals of healthy larynx, 51 signals of paralysis, 43 signals of Reinke's edema, and 18 signals of nodules is taken into account.

Results

The classification performance for the selected features is shown in Table 1. The accuracy refers to the correctly classified cases, including all analyzed signals. The specificity means the correct classification of health cases (true negative) and the sensitivity corresponds to the pathological voices correctly classified (true positive).

Table1– Classification performance obtained withPSO algorithm and MLP.

Classification Cases	Accuracy (%)	Specificity (%)	Sensitivity (%)
HTYxPTL	88±5	87±4	63±10
HTYxPRL	91±4	74±9	93±5
HTLxEDM	86±4	62±8	88±4
HTYxNDL	91±3	50±10	96±2

Conclusion

The use of wavelet transform and Haralick features may be considered as an applicable tool to texture analysis of recurrence plots in order to identify laryngeal pathologies. For validating the technique, practical tests are necessary.

References

[1] Costa WCA. Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas [thesis]. Campina Grande, Universidade Federal de Campina Grande, 2012.

[2] Marwan N.Encounters With Neighbours - Current Developments of Concepts Based on Recurrence Plots and Their Applications [thesis]. Potsdam, University of Potsdam, 2003.

[3] Haralick RM, Dinstein I, Shanmugam, K. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 1973: (3): 610-621.

[4] Mallat, SGA. Wavelet tour of signal processing. San Diego: Academic Press, 1999.

[5] EberhartR and Kennedy J. A new optimizer using article swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya,Japan, 1995, 39–43.

[6] Kay Elemetrics, Kay Elemetrics Corp. Disordered Voice Database, Model 4337, 1994.

Address for correspondence

Taciana Araújo de Souza. Coordenação de pós-graduação em Engenharia Elétrica. Av. Primeiro de Maio, 720, Jaguaribe, João Pessoa – PB, CEP: 58015-430; tacianafisica@gmail.com.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1048

On The Correlation Between Geo-Referenced Clinical Data And Remotely Sensed Air Pollution Maps

Arianna Dagliati[a], Andrea Marinoni[b], Carlo Cerra[c], Paolo Gamba[b], Riccardo Bellazzi[a]

[a] *Laboratory for Biomedical Informatics, University of Pavia, Italy*
[b] *Telecommunications and Remote Sensing Laboratory, University of Pavia, Italy*
[c] *Department of Purchasing and Control, ASL Pavia, Italy*

Abstract

This work presents an analysis framework enabling the integration of a clinical-administrative dataset of Type 2 Diabetes (T2D) patients with environmental information derived from air quality maps acquired from remote sensing data. The research has been performed within the EU project MOSAIC, which gathers T2D patients' data coming from Fondazione S.Maugeri (FSM) hospital and the Pavia local health care agency (ASL). The proposed analysis is aimed to highlight the complexity of the domain, showing the different perspectives that can be adopted when applying a data-driven approach to large variety of temporal, geo-localized data. We investigated a set of 899 patients, located in the Pavia area, and detected several patterns depicting how clinical facts and air pollution variations may be related.

Keywords:

Geo-referenced data; Environmental factors; Type 2 Diabetes; Air pollution maps.

Introduction

The environment's involvement in health issues has been defined as the Exposome [1]. One of the main advantages of the available data sets is that, thanks to the information derived from the local healthcare agency, patients' addresses are associated to municipality codes. Precisely locating patients enables the investigation of their interactions with the territory. It has also been stated that the combination of wide-ranging data sources requires innovative analysis approaches from methods and data management point of view [2]. In our case a further step in the study was triggered by merging biomedical informatics with telecommunications and remote sensing domain expertise. We exploited geo-referenced clinical data to understand the effects of being exposed with air pollutants thanks to air quality maps acquired from remotely sensed data.

Materials and Methods

The joint analysis of air quality maps and geo-localized clinical events gave the possibility to observe multiple patterns varying over time. Comparing these trends provided some insight about the evolution of severe events, which are those requiring hospitalization.

In order to achieve air quality maps, the data acquired by Landsat L8 mission have been considered. Landsat 8 satellite was launched in February of 2013. The satellite collects images of the Earth with a 16-day repeat cycle, referenced to the Worldwide Reference System-2. Thanks to the assumption that pollution plays a key-role in the thermal pattern of a remotely sensed scene [3], the correlation between the presence of black particulate and the recorded temperature can be thoroughly characterized. In order to match the region of the remotely sensed data with the Pavia area, we implemented spectral analysis taking into account the overall pattern of the raw counts thermal signal as a function of black particulate concentration. For each pixel in every temporal series, a polynomial fitting model has been implemented to estimate the air quality of the scene [3]. Air quality has been quantized on five levels over the black particulate concentration estimate.

While observing that, during the analyzed time window, the number of hospitalizations in two considered areas follows inverse trend during the coldest season, we were able to describing those events per-se. Although, thanks to additional source of data, we can also infer that there are other external factors entailing the observed facts, so to enable the discovery of additional explanations of clinical phenomenon.

Results

Through the information acquired from air-quality maps we observed that, in different areas, the increased number of hospitalization was related to different value of air pollutants. In particular we focused on the area named Lomellina during the February month. In this period in-hospital events account for the 31.58% of the entire observations, which is the highest value compared with other areas; within the same time window and area it was possible to detect worsened condition from air quality maps (Particulate Matter values increased from 51/71 mg/m^3 in December to 76/100 mg/m^3 in February).

Conclusions

Merging heterogeneous information coming from various scientific disciplines is the trigger to inspect complex evolution patterns of diseases. Our findings provided meaningful hints to investigate further how clinical acute events may be explained by exogenous factors.

References

[1] Wild, CP. "Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology." Cancer Epidemiol Biomarkers August 2005: 14: 1847.

[2] Martin S. "Exposome informatics: considerations for the design of future biomedical research information systems." J Am Med Inform Assoc. 2014: May-Jun; 21(3):386-90.

[3] L. Wald . "Observing air quality over the city of nantes by means of landsat thermal infrared data." Int.J. of Remote Sensing 1999: 20(5):947–959.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1049

Integrating data from multiple sources for data completeness in a web-based registry for pediatric renal transplantation - the CERTAIN Registry

Lennart Köster[a], Kai Krupka[b], Britta Höcker[b], Axel Rahmel[ce], Undine Samuel[c], Wouter Zanen[c], Gerhard Opelz[d], Caner Süsal[d], Bernd Döhler[d], Lukasz Plotnicki[b], Christian D. Kohl[a], Petra Knaup[a], Burkhard Tönshoff[b]

[a] *Institute of Medical Biometry and Informatics, University Hospital of Heidelberg, Germany*
[b] *Department of Pediatrics I, University Children's Hospital, Heidelberg, Germany*
[c] *Stichting Eurotransplant International Foundation, Leiden, The Netherlands*
[d] *Department of Transplantation Immunology, University Hospital of Heidelberg, Germany*
[e] *German Organ Transplantation Foundation, Frankfurt, Germany*

Abstract

Patient registries are a useful tool to measure outcomes and compare the effectiveness of therapies in a specific patient population. High data quality and completeness are therefore advantageous for registry analysis. Data integration from multiple sources may increase completeness of the data. The pediatric renal transplantation registry CERTAIN identified Eurotransplant (ET) and the Collaborative Transplant Study (CTS) as possible partners for data exchange. Import and export interfaces with CTS and ET were implemented. All parties reached their projected goals and benefit from the exchange.

Keywords:

Patient registry; data integration; interfaces; pediatric-kidney transplantation.

Introduction

Patient registries collect data about the patients' course of disease or therapy. Registries rely on good data quality and completeness as the observations made from the available data only pertain to the observed population and are not generalizable [1]. To increase data completeness, the pediatric renal transplantation registry CERTAIN [2] decided to exchange data with ET and CTS to increase data completeness.

Materials and Methods

The CERTAIN Registry is an in-house developed registry written in Java with a web-client based on Google Web Toolkit and an SQL database. Interfaces with other data sources and organizations are implemented in-house. Currently 51 centers from 10 countries contribute to the registry.

Results

Import of data from CTS provided CERTAIN with new patients and long-term follow-up data. Export to CTS prevents double-entry of CERTAIN-data into CTS. High quality donor and recipient data could be obtained through the continuous data exchange with ET. Exporting follow-up data to ET provides a larger data basis for future evidence-based improvement of organ allocation. Up to now, data from 285 patients have been imported from CTS. ET has received follow-up data for 1096 visits of 193 patients. CERTAIN received notification of 436 transplantations from ET.

Discussion

Establishing exchanges with ET and CTS was a major step towards data completeness while easing data entry for participating centres. Interface-implementation required considerable coordination and effort from all parties. Privacy concerns proved to be the biggest challenge to be solved. Reconciliation of data sets proved to be time-consuming.

Conclusion

We established the communication and data exchange of the CERTAIN Registry with CTS and ET respectively. Both interfaces increased the completeness of CERTAIN's data set. The data exchange has proven to be beneficiary for the CERTAIN Registry and the participating organizations. Data exchange with other registries may further benefit all parties.

References

[1] Lim TO, Goh A, Lim, YN and Morad Z. Review article: Use of renal registry data for research, health-care planning and quality improvement: What can we learn from registry data in the Asia–Pacific region?. Nephrology 2008: 13 (8): pp. 745-52.

[2] Plotnicki L, Kohl CD, Höcker B, Krupka K, Rahmel A, Pape L, Hoyer P, Marks SD, Webb NJA, Söylemezoglu O, Topaloglu R, Szabo AJ, Seeman T, Cornelissen EAM, Knops N, Grenda R and Tönshoff B. The CERTAIN registry: a novel, web-based registry and research platform for pediatric renal transplantation in Europe. Transplant Proc. 2013; 45 (4): pp. 1414-7.

Address for correspondence

Lennart Köster, Institute of Medical Biometry and Informatics, University Hospital of Heidelberg, lennart.koester@med.uni-heidelberg.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1050

Development of Unified Lab Test Result Master for Multiple Facilities

Naoto Kume[a], Kenji Suzuki[b], Shinji Kobayashi[a], Kenji Araki[c] and Hiroyuki Yoshihara[c]

[a] EHR research unit, Kyoto University, [b] SRL Inc, [c] Miyazaki University Hospital

Abstract

A clinical study requires massive amounts of of lab test data, especially for rare diseases. Before creating a protocol, the hypothesis if the protocol will work with enough amount of patients' dataset has to be proved. However, a single facility, such as a university hospital, often faces a lack of number of patients for specific target diseases. Even if collecting datasets from several facilities, there is no active master table that can merge lab test results between the facility datasets. Therefore, the authors develop a unified lab test result master. Because test master standards such as JLAC10 and LOINC are provided from a viewpoint of academic classification of laboratory medicine, the classification does not fit clinical classification, which doctors understand with a mind-set of establishing a clinical study protocol. The authors establish a method to unify masters using an active lab test result master from two university hospitals.

Keywords:

EHR; Lab test master; JLAC10.

Introduction

For developing a clinical study protocol, the expected number of patients is important to decide if the study is actually carried out. In a single facility, such as a university hospital, it is difficult to assume how many patients are in a region. Therefore, researchers have to assume a high burden for case finding by communicating with multiple facilities. Even if all test results are provided from participant facilities, merging results is another huge burden before analysis. Therefore, a unified lab test master is required to compare datasets between different facilities.

Beforehand, the authors tried to use a standard master JLAC10, which is provided by the Japanese Society of Laboratory Medicine. The results reflect that the master is classified from a viewpoint of academic classification of lab tests so that it was difficult to merge test results between facilitites. At the trial, four university hospital's masters, with about three to five thousand lines, are matched. Only a hundred items can be described by the unified master.

Therefore, the authors propose a master table merging method to establish a unified master from a viewpoint of a doctor's mind, which is adopted when a clinical study is designed.

Methods

To decrease the burden of master table matching, the methods proceeds with three steps. Eventually, the proposed unified master code consists of a combination of high-order digits and two original sub-codes. For the first step, an actual lab test result dataset is analyzed to select active items. Since an active lab test master is always maintained in each facility, some items are active and some are not. Two years lab test results provide active test items, traditional items, and inactive items. For the second step, an accounting information table and a test order item table are provided. The test order item contains in many cases multiple test result items. In some cases, order items contains a single test result item. Therefore, a unified master code high-order number is given by an accounting information table code, which is provided by the Ministry of Health, Labour and Welfare in Japan. Since in Japan the accounting code is unified, every facility has the same code. The rough classification of lab result items has been analyzed with the accounting code. Figure 1 depicts the unified master code construction.

For the third step, matchings of each item from unified code and facility code are analyzed by manually checking item names. Sub-code 1 classifies specimen and dose amount. Sub-code 2 classifies details such as methods, units and applied standards.

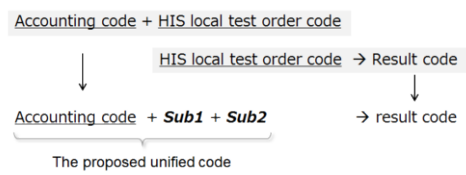


Figure 1. Unified code structure

Results

A beta version of a unified code is generated by two university master codes. The total amount of unified codes covers around 800 result items. Some original facility tests and subcontract lab test results are not yet stored in the unified code.

Conclusion

A unified lab test result master is proposed for clinical study case findings. The combination code of accounting code and sub-classification code provides the unified master code.

References

[1] Kume N, Manabe S, Matsumura Y, Araki K, Takabayashi K, Shima Y, Okamoto K, Takemura T, Kuroda T, Yoshihara H, Distributed Clinical Study Database for Traversal Search of the Accounting Information and Laboratory Test Results, IEEE EMBC Short Papers, no.1686, 2013.

Address for correspondence

Naoto Kume, kume@kuhp.kyoto-u.ac.jp, EHR Research Unit, Department of Social Informatics, Graduate School of Informatics, Kyoto University. 506, Kyoto Research Park Dept. 9, Awatacho 91, Chudoji, Shimogyo-ku, Kyoto-city, 600-8815, Japan.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1051

Semantic Web Ontology and Data Integration: a Case Study in Aiding Psychiatric Drug Repurposing

Chen Liang, Jingchun Sun, Cui Tao

School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

Abstract

There remain significant difficulties selecting probable candidate drugs from existing databases. We describe an ontology-oriented approach to represent the nexus between genes, drugs, phenotypes, symptoms, and diseases from multiple information sources. We also report a case study in which we attempted to explore candidate drugs effective for bipolar disorder and epilepsy. We constructed an ontology incorporating knowledge between the two diseases and performed semantic reasoning tasks with the ontology. The results suggested 48 candidate drugs that hold promise for further breakthrough. The evaluation demonstrated the validity our approach. Our approach prioritizes the candidate drugs that have potential associations among genes, phenotypes and symptoms, and thus facilitates the data integration and drug repurposing in psychiatric disorders.

Keywords:

Drug Repurposing; Ontology; Data Integration.

Introduction

Despite the remarkable progress in fundamental research and medications, efforts toward the discovery of new drugs for psychiatric disorders have been relatively unsuccessful compared to drug discovery for other diseases. There is an imperative for discovering novel solutions. Researchers have attempted to develop new solutions including drug repurposing [1], which aims to develop new uses for existing or abandoned drugs [2]. Drug repurposing methods encompass a wide spectrum ranging from traditional screening methods and animal models to computational methods [3]. Computational methods hold promise for facilitating the identification of candidate drugs with the advantage of time efficiency. We proposed a framework built under the W3C Standard of Web Ontology Language (OWL). We also described a use case with a semantic reasoning task using the ontology to explore candidate drugs activating for bipolar disorder and epilepsy.

Methods

We defined five top-level classes (disease, symptom, phenotype, gene, and drug), along with the corresponding OWL object properties to represent the metaontology. The detailed ontology includes OWL individuals that were defined under the corresponding classes to model real-world entities. In our ontology, these instances were imported primarily from open data repositories, including PharmGKB, Human Phenotype Ontology, and Online Mendelian Inheritance in Man. We encoded symptom-associated information for the ontology since there is no existing data repository providing such information. In the use case, we generated Descriptive Logic (DL) Rules for delivering reasoning results. For example, one DL rule to find a new path between drugs and diseases can be defined as *"DrugRelatedToGene some (associatedWithPhenotype some (associatedWithSymptom some ((associatedWithDisease some Bipolar_disorder) and (associatedWithDisease some Epilepsy))))"*. A candidate drug inferred by this DL must be a drug that associates with symptoms that appear in Bipolar_disorder and Epilepsy. The associations between drug and symptom were inferred through drug-gene-phenotype-symptom. To validate the inferred candidate drugs and ontology effectiveness, we used data from PharmGKB, ClinicalTrial.gov, and PubMed for the evaluation.

Results

The ontology inferred 48 drugs relevant to the two diseases. We further evaluated whether symptoms contain important information in guiding drug selection. The result shows that one out of the 48 drugs has direct disease-gene-drug associations. Eleven drugs have been investigated in the clinical trials on both bipolar disorder and epilepsy. Eighteen drugs from our queried results that revealed joint occurrences of the drug name and the two disease names in PubMed, among which only four drugs revealed weak evidence (returned the occurrence of one). Eleven drugs were identified from both ClinicalTrial.org and PubMed.

References

[1] Iorio F, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 2010: 107(33): 14621-14626.

[2] Carley DW. Drug repurposing: identify, develop and commercialize new uses for existing or abandoned drugs. Part I. IDrugs: the investigational drugs journal, 2005: 8(4): 306.

[3] Dudley JT, Deshpande T, & Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. Briefings in bioinformatics, 2011.

Address for correspondence

Cui Tao, PhD
7000 Fannin, UCT 600, Houston, TX, 77030, USA
Cui.Tao@uth.tmc.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1052

Personalised Medicine Possible With Real-Time Integration of Genomic and Clinical Data To Inform Clinical Decision-Making

Fernando Martin-Sanchez[a], Maureen Turner[a], Alice Johnstone[a], Leon Heffer[a], Naomi Rafael[a], Advisory Group[b], Tim Bakker[b], Natalie Thorne[b], Ivan Macciocca[b], Clara Gaff[b]

[a] *BioGrid Australia, Melbourne, VIC, AUSTRALIA*
[b] *Melbourne Genomics Health Alliance, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, AUSTRALIA*

Abstract

Despite widespread use of genomic sequencing in research, there are gaps in our understanding of the performance and provision of genomic sequencing in clinical practice. The Melbourne Genomics Health Alliance (the Alliance), has been established to determine the feasibility, performance and impact of using genomic sequencing as a diagnostic tool.

The Alliance has partnered with BioGrid Australia to enable the linkage of genomic sequencing, clinical treatment and outcome data for this project.

This integrated dataset of genetic, clinical and patient sourced information will be used by the Alliance to evaluate the potential diagnostic value of genomic sequencing in routine clinical practice. This project will allow the Alliance to provide recommendations to facilitate the integration of genomic sequencing into clinical practice to enable personalised disease treatment.

Keywords:

Translational Medical Research; Process Assessment (Health Care); Data Linkage; Genomics; Individualized Medicine.

Introduction

Over 14,000 genes are known to cause or contribute to disease [1] and over 2,000 diagnostic genetic tests are now available in clinical practice [2]. Despite widespread use of genomic sequencing in research, there are gaps in our understanding of the performance and provision of genomic sequencing for clinical use. The Melbourne Genomics Health Alliance (the "Alliance") was launched as a collaboration between seven Melbourne-based research and clinical organisations. The Alliance has been established to determine the feasibility, performance and impact of using genomic sequencing as a diagnostic tool with the objective of demonstrating that personalised medicine through targeted genomic analysis is possible.

Methods

To enable the linkage of genomic sequencing, clinical treatment and outcome data, the Alliance has partnered with BioGrid Australia for this project. BioGrid Australia operates a secure federated data-sharing platform that enables real-time integration of record-level data across institutions and jurisdictions. This platform provides ethical access to data while protecting both privacy of the participants and the intellectual property of the contributing organizations. Importantly, the platform has the capability to provide an integrated view for the treating clinician as well as project wide reporting.

Results

This project has created a unique integrated dataset of diverse genetic, clinical and patient survey information to evaluate the potential diagnostic value of genomic sequencing in routine clinical practice. As the project continues into its next phase, the Alliance will build on these initial results to gain a better understanding of the process requirements to use genomic sequence data in clinical care.

Discussion

As the use of genomic sequencing information in clinical care is limited, the Alliance has created a novel way to facilitate the integration of genomic sequencing into clinical practice. The results of the demonstration project will be used by the Alliance to provide process recommendations for subsequent implementation to improve both the impact and the performance of genomic sequence information in clinical practice and personalised care.

Conclusion

These translational analyses provide the groundwork for identifying the processes required for further exploration of targeted personalised disease treatment in clinical practice.

References

[1] Online Mendelian Inheritance in Man, from: www.omim.org/statistics/entry, accessed 17/10/13
[2] Centre for Disease Control, from: http://www.cdc.gov/genomics/gtesting/, accessed 18/10/13

Address for correspondence

enquiries@biogrid.org.au
enquiries@melbournegenomics.org.au

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1053

Integration of Disease Specific Clinical and Genomics Datasets using I2B2 Framework

Mohyuddin[a]

[a] *King Abdullah International Medical Research Center (KAIMRC), National Guard Health Affairs (NGHA), Riyadh, Saudi Arabia.*

Abstract

The availability of a patient's genomic profile along with the clinical profile for providing individualized care and treatment is paving the road for a new era of personalized medicine, and is an important area of focus in current biomedical research. One of the prominent and globally implemented solutions for clinical and genomics data integration in biomedical research is an NIH funded NCBC initiative - Informatics for Integrating Biology and the Beside (I2B2). This paper presents the development of a pilot prototype for integrating patient's clinical and genomics datasets using open source and scalable I2B2 Framework. It focuses on disease specific clinical data and genomic variants, when combined together can be used for informed decision making in clinical practices by healthcare professionals and for further investigations by biomedical researchers. The research was carried out using a case study of King Abdullah International Medical Research Center (KAIMRC) in collaboration with King Fahad National Guard Hospital (KFNGH).

Keywords:

Integrated Health Care Systems; Cohort Analysis; Information Retrieval; Dataset; Biomedical Research

Introduction

Since the completion of Human Genome Project, biomedical literature has shown abundance of various approaches for data integration, such as federated databases, data warehouses and data management systems; but most of them focus only on genomics data integration from heterogeneous resources [1, 2]. I2B2 provides an environment where users can perform queries and cohort analysis by integrating clinical and biomedical data together, that can be used for discovery research and targeted therapies for individual patients having genetic diseases [3]. For the proof of concept, we developed a pilot prototype to evaluate feasibility of using I2B2 for integrating anonymized patients' clinical and genomics datasets.

Methods

The sample clinical dataset included patients' demographics and basic disease profiles, such as gender and age information, disease type, stage and level. The sample genomic variations data, prepared by studying DECIPHER project [4], included number of genes, inheritance levels and variant locations containing chromosome number, start location and end location. Both datasets were further customized for mapping and interoperability with I2B2 star schema and ontology.

Results

I2B2 data warehouse (hosting our clinical and genomics datasets) and data management services (core server-side cells - the hive and client-side applications - web client and workbench) were installed and configured in Windows and Linux Platforms. In Windows, we deployed I2B2 image using virtualization software - VMware Player; whereas in Linux - we used terminal along with I2B2 Wizard [5] to install and configure the requisite I2B2 softwares. After datasets' customization, IDRT Import Tool [5] was used for importing both datasets in I2B2 data warehouse. Finally, the web client was used for data query, cohort analysis and information visualization.

Figure 1-Clinical and Genomics Data Integration using I2B2

Conclusion

We found that I2B2 installation, configuration and data import process is complex, and also multiple data formats support is limited. The successful implementation of current pilot prototype has laid foundations of our future work where we will target the integration of local patients' disease and genetic registries based on stakeholders' requirements using I2B2.

References

[1]Louie, B., et al., Data integration and genomic medicine. J Biomed Inform, 2007. 40(1): p. 5-16.

[2]Zhang, Z., et al., Data Integration in Bioinformatics: Current Efforts and Challenges, in Bioinformatics - Trends and Methodologies, D.M.A. Mahdavi, Editor. 2011.

[3]NCBC. Informatics for Integrating Biology and the Bedside (I2B2). 2014; Available from: https://www.i2b2.org/.

[4]Bragin, E., et al., DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res, 2014. 42(Database issue).

[5]IMISE, U.L. Integrated Data Repository Toolkit (IDRT). 2014; Available from: http://idrt.imise.uni-leipzig.de/

Address for correspondence

Mohyuddin, KAIMRC, NGHA, KAMC, P.O.Box. 22490, Riyadh 11426-1515, Saudi Arabia; Email: mrmohy@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1054

Bridging the Gap from Bench to Bedside – An Informatics Infrastructure for Integrating Clinical, Genomics and Environmental Data (ICGED)

Mohyuddin

King Abdullah International Medical Research Center (KAIMRC), National Guard Health Affairs (NGHA), Riyadh, Saudi Arabia.

Abstract

The abundance of heterogeneous biomedical data from a variety of sources demands the development of strategies to address data integration and management issues, so that the data can be used effectively in clinical practices and biomedical research. This research presents an Informatics Infrastructure for Integrating Clinical, Genomics and Environmental Data (ICGED) and provides a roadmap that envisions utilizing the clinical and biomedical resources in our case study. This work describes a data integration approach, proposed by ICGED, with a two-fold purpose: personalized medicine and biomedical data storage and sharing platform. It describes our experiences integrating disease specific clinical and genomics datasets with Data Integration and Analysis Tools (DIAT)—using Informatics for Integrating Biology and the Bedside, and discusses work in progress and future work for extending DIAT, and the development of Risk Assessment and Prediction Tools, Clinical Decision Support Systems and a Bioinformatics Data Warehouse.

Keywords:

Personalized Medicine; Integrated Health Care Systems; Information Sharing; Risk Assessment; Decision Making.

Introduction

Personalized Medicine provides individualized treatment based upon a patient's unique clinical, genomic and environmental profile, and likelihood to a drug/therapy response, by using precise risk assessment and clinical decision support tools [1]. This research investigates the development and utilization of such tools in the resource structure of our case study of the King Abdullah International Medical Research Center and the King Fahad National Guard Hospital. Initial requirements analysis showed a need for using translational research and biomedical data infrastructure.

Methods

This research proposes an Informatics Infrastructure for Integrating Patient's Clinical, Genomics and Environmental Data (ICGED), which provides a centralized platform for data deposition and analytics, information access and sharing, and knowledge extraction. This platform will allow healthcare staff to find relevant information for individualized patient treatment, and biomedical researchers to utilize existing information for further research. Figure 1 shows the ICGED roadmap for our on-going and future research work.

Figure 1 – ICGED Roadmap

Results

For proof of concept, a pilot prototype was developed as a part of Data Integration and Analysis Tools (DIAT) by integrating sample clinical and genomics datasets using Informatics for Integrating Biology and the Bedside (i2b2) [2]. The i2b2 data warehouse and data management services were configured using Windows (XP) and Linux (Ubuntu 12.04). VMware Player and i2b2 Wizard [3] were used for environment setup, datasets were customized according to star schema, and data mapping and import was done using the Integrated Data Repository Toolkit [3]. Finally, cohort analysis and visualization was done using the i2b2 Web Client [2].

Conclusion

This work presented ICGED Infrastructure, roadmap and pilot prototype as a DIAT tool. In the future, we plan to extend DIAT using disease and genetic registries, develop RAPT using a machine learning model, CDSS using I2B2 plug-ins [2], and BDW using existing standards [4]. Some important challenges of personalized medicine, such as lack of experts/training, and social, ethical and legal aspects of genetic information use must be addressed in future research work.

References

[1] Zhang, P., et al., Towards personalized medicine: leveraging apatient similarity and drug similarity analytics. AMIA Jt Summits Transl Sci Proc, 2014: p. 132-6.

[2] NCBC. Informatics for Integrating Biology and the Bedside (i2b2). 2014; Available from: **https://www.i2b2.org/**.

[3] IMISE, U.L. Integrated Data Repository Toolkit (IDRT). 2014; Available from: **http://idrt.imise.uni-leipzig.de/**.

[4] Fernandez-Suarez, X.M. and M.Y. Galperin, The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. Nucleic Acids Research, 2012.

Address for correspondence

Mohyuddin, KAIMRC, NGHA, KAMC, P.O.Box. 22490, Riyadh 11426-1515, Saudi Arabia; Email: mrmohy@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1055

Designing an Innovative Data Architecture for the Los Angeles Data Resource (LADR)

Sukrit Mukherjee[a], Robert A. Jenders[a,b], Sebastien Delta[a]

[a] *Center for Biomedical Informatics, Charles R. Drew University of Medicine and Science, Los Angeles, CA, USA*
[b] *Department of Medicine, University of California, Los Angeles, CA, USA*

Abstract

The Los Angeles Data Resource (LADR) is a joint project of major Los Angeles health care provider organizations. The LADR helps clinical investigators to explore the size of potential research study cohorts using operational clinical data across all participating institutions. The Charles R. Drew University of Medicine and Science (CDU) LADR team sought to develop an innovative data architecture that would aggregate de-identified clinical data from safety-net providers in the community into CDU LADR node. This in turn would be federated with the other nodes of LADR for a shared view in a way that was never available before. This led to a self-service system to assess patients matching study criteria at each medical center and to search patients by demographics, ICD-9 codes, lab results and medications.

Keywords:
Data Architecture, Electronic Health Record, Data Staging, Data Transformation, Grid Node Database, Federated Query.

Introduction

LADR is a joint project of major Los Angeles health care provider organizations. It facilitates clinical investigators to explore the size of potential research study cohorts across each participating institution using de-identified clinical data captured during customary clinical operations under the Clinical and Translational Science Institute (CTSI) program. CDU is a member of the CTSI but unlike the other LADR members that lacks clinical operations of its own, it has a rich history of partnerships with safety-net ambulatory providers serving the poor and minority neighborhoods of South Los Angeles. The CDU LADR team sought to develop a new data architecture that would aggregate data from disparate health information systems in the community into the CDU LADR node.

Methods

Using information provided by 'L.A. Net' - a primary care practice-based research network (PBRN) in the region partnering with more than 165 unique practice sites in Southern California - as a case study, we developed an extensible data architecture to obtain data from small health care organizations for LADR. L.A. Net had its own aggregator resource, the Reusable OMOP-SAFTINet Interface Transformation Adaptor (ROSITA) that supports the consumption, transformation and loading of clinical and administrative data from its partners' electronic health record systems (EHRs) and payer claims data into L.A. Net's grid node database. We designed an architecture to aggregate data from ROSITA and other individual EHRs into CDU LADR application server via a web service method of communication over the Internet.

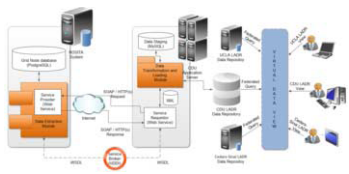


Figure 1 – The architecture for CDU LADR node

Results

The service requester module located on our server would request data, whereas the service provider module designed by us and placed on the ROSITA system would pull the requested operational clinical data from the grid node database, remove 18 specific identifiers (Safe Harbor Method) to de-identify the data thus complying with USA regulation regarding limited datasets and extract the de-identified data. Next, the service requester would receive data over SOAP or HTTPS protocol as XML to be loaded first into the staging database and later into the CDU's LADR Data warehouse. This XML file would be validated by the service requester for correct/expected values using an XSD file. A UDDI (Universal Description, Discovery and Integration) process would be defined to identify which software system should be contacted for which type of data. Thus, the CDU LADR node will have its own data repository and eventually be able to federate data with other participating institutions for a shared view.

Conclusion

We created a standards-based, flexible architecture to federate data with other LADR nodes for a shared view leading to a self service system. This system assessed patients matching study criteria at participating medical centers to search by demographics, ICD-9 codes, lab results, medications and to generate enrollment tables for clinical and translational research.

Acknowledgments

This work was supported in part by US National Institutes of Health's NIMHD grant number U54 MD007598.

Address for correspondence

Sukrit Mukherjee, MS, MSE, PhD
Center for Biomedical Informatics, Charles R. Drew University of Medicine and Science, 1731 E 120th St, Building N, Los Angeles, CA 90059 USA. Email: sukritmukherjee@cdrewu.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1056

Curating and Integrating Data from Multiple Sources to Support Healthcare Analytics

Kenney Ng[a], Chris Kakkanatt[b], Michael Benigno[b], Clay Thompson[b], Margaret Jackson[b], Amos Cahan[a], Xinxin Zhu[a], Ping Zhang[a], Paul Huang[c]

[a] *IBM T.J. Watson Research Center, Yorktown Heights, NY, USA*
[b] *Pfizer Inc., New York, NY, USA*
[c] *Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*

Abstract

As the volume and variety of healthcare related data continues to grow, the analysis and use of this data will increasingly depend on the ability to appropriately collect, curate and integrate disparate data from many different sources. We describe our approach to and highlight our experiences with the development of a robust data collection, curation and integration infrastructure that supports healthcare analytics. This system has been successfully applied to the processing of a variety of data types including clinical data from electronic health records and observational studies, genomic data, microbiomic data, self-reported data from surveys and self-tracked data from wearable devices from over 600 subjects. The curated data is currently being used to support healthcare analytic applications such as data visualization, patient stratification and predictive modeling.

Keywords:
Data Collection; Data Curation; Automatic Data Processing.

Introduction

The volume and variety of healthcare related data continues to grow, spurred on by the increasing adoption and use of electronic health records (EHRs), the explosion of omics data and the proliferation of actigraphy data from wearable self-tracking devices and mobile applications [1]. The use of this data for healthcare analytics such as data visualization, patient stratification, predictive modeling, personalized medicine and drug discovery will increasingly depend on the ability to appropriately collect, curate and integrate disparate data from many different sources [2]. The linking of many different types of data will allow the analysis and modeling of complex multi-dimensional interactions which can enable deeper insights.

Materials and Methods

Our data consist of the following diverse types and sources for over 600 subjects: (1) clinical data from EHRs and observational studies: demographics, family history, labs, vitals, diagnoses, medications; (2) genomic data: 500K single nucleotide polymorphisms (SNPs); (3) microbiomic data: abundances of gut microbial taxa; (4) self-reported data from surveys: behavioral, lifestyle, diet; and (5) self-tracked data from wearables: activity, sleep, diet.

To process these data, we developed a robust data collection, curation and integration infrastructure composed of the following stages: (1) **Data Collection**: gathering raw data from different sources. (2) **Data Understanding**: characterizing the data fields, types, and values. (3) **Data Validation**: checking the data against known quantitative relationships and expected values and for consistency across data types. Examples include checking lab values against standard ranges and using system physiology models to identify potential outliers in reported caloric intake data. (4) **Data Cleaning**: cleaning, normalizing and mapping data values and tagging the data with confidence indicators. Examples include using natural language processing to extract and normalize noisy medication values and map them to standard ontologies and flagging suspicious self-reported and self-tracked data based on compliance definitions. (5) **Data Integration**: merging data from different sources, resolving ambiguities, deduplication, normalizing units and dates and linking variables to the same subject. (6) **Data Enrichment**: creating new variables from the original data that are potentially more informative. Examples include the computation of genetic risk scores from the SNPs and the computation of scores from survey responses.

The data are then stored in a database using an n-tuple structure: subject identifier, feature identifier, feature value, confidence and event date that accomodates heterogenous data and supports the data retrieval needs of analytic applications.

Results

We found that a semi-automated approach, where most of the processing is automated but unhandled issues and errors can be flagged for human intervention, was important. Because some data were human reported and entered, they contained errors. As a result, it was important to have validation methods to check, clean, normalize and map the data when possible. In addition, we tagged the data instances with confidence indicators to allow analytic applications to select appropriate data subsets for their own use. When possible, we leveraged the different data types to perform cross type consistency checking. Incremental update capability was also needed to support the addition of new data to the database. Finally, we used an iterative development and refinement process in order to accommodate enhancements resulting from new data types, new instances of existing data types and feedback from the analytic applications that consume the curated data.

Conclusions

A robust infrastructure was successfully developed and used to collect, curate and integrate EHR, genomic, microbiomic, self-reported and self-tracked data to support healthcare analytics. Future work will enhance and extend the system to handle additional types of data including metabolomics, continuous biomarker data streams and medical imaging data.

References

[1] Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D.. Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. Biomed Res Int. 2014.

[2] Stonebraker M, Bruckner D, Ilyas IF, Beskales G, Cherniack M, Zdonik SB, Pagan A, Xu S. Data Curation at Scale: The Data Tamer System. CIDR 2013.

Address for correspondence:

Kenney Ng (kenney.ng@us.ibm.com)
IBM Research, 1 Rogers Street, Cambridge MA 02142, USA.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1057

Integrated Database And Knowledge Base For Genomic Prospective Cohort Study In Tohoku Medical Megabank Toward Personalized Prevention And Medicine

Soichi Ogishima[a], Takako Takai[a], Kazuro Shimokawa[a], Satoshi Nagaie[a], Hiroshi Tanaka[a], Jun Nakaya[a,b]

[a] *Dept. of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan*
[b] *Dept. of Medical Informatics, Tohoku University School of Medicine, Miyagi, Japan*

Abstract

The Tohoku Medical Megabank project is a national project to revitalization of the disaster area in the Tohoku region by the Great East Japan Earthquake, and have conducted large-scale prospective genome-cohort study. Along with prospective genome-cohort study, we have developed integrated database and knowledge base which will be key database for realizing personalized prevention and medicine.

Keywords:

Tohoku Medical Megabank project, prospective genome-cohort study, integrated database and knowledge base

Introduction

To revitalize medical care and realize personalized prevention and medicine in the disaster area of Great East Japan Earthquake, the Tohoku Medical Megabank Organization has been established in February 2012. It aims at becoming a center for the reconstruction of the entire Tohoku region by conducting prospective genome-cohort study and developing the large-scale genome biobank toward personalized prevention and medicine.

In our prospective cohort study, we will recruit 150,000 people at Tohoku university, satellites, health clinics, and Iwate medical university, and collect (1) biospecimen (blood, urine) (2) questionnaire, (3) physical measurement as baseline and follow-up investigation. As for pathogenesis investigation follow-up, we will collect clinical data from electronic health records provided by Miyagi Medical and Welfare Information Network (MMWIN). MMWIN is expected to develop regional electronic network of medical records of most hospitals in Miyagi prefecture. Collected data will be de-identified and stored in a database of Tohoku Medical Megabank. Researchers in universities, institutes, and companies can apply withdrawal of biospecimen and data.

Materials and Methods

Integrated database provides large genome-cohort study data. It integrates various and vast amount of data. It is too huge to to grasp characteristic of data. Therefore, integrated database not only provides integration of data but also provides statistical significance and related knowledge in cooperation with knowledge base. Knowledge base provides existing knowledge of personalized prevention and medicine with NLP support for extraction of knowledge from literature, and novel knowledge by large-scale correlation analysis.

Results

Integrated database is a database for integration of genomic (omics), specimen, baseline & follow-up and clinical data. It is a relational database implemented by Oracle and MySQL RDMS. On the other hand, knowledge base is a database for integration of existing and novel knowledge by large-scale correlation analysis on integrated database. It is a RDF store implemented by Virtuoso. Our database is is too huge to grasp characteristic of data. Therefore, integrated database not only provides integration of data but also provides statistical significance and related knowledge in cooperation with knowledge base.

Statistical significance of selected data is provided in integrated database to characterize selected cohort; e.g., selected population has statistically significant bias in sex and alcohol intake. This characterization is expected to lead to hypothesis. Related knowledge of selected data is also provided in integrated database in cooperated with knowledge base; e.g., breast cancer is known to have a highly linked and causative SNPs. This provision of related knowledge is expected to accelerate studies.

Our project collected biospecimen and data provided by living people. Thus, data security and protection is very important. Our databases will be available with keeping security under our security policies (personal, very strong, strong and standard security policies).

Conclusion

The Tohoku Medical Megabank project is a national project conducting large-scale prospective genome-cohort study. We have developed integrated database and knowledge base which will be key database for realizing personalized prevention and medicine.

Acknowledgments

This work was supported by the MEXT Tohoku Medical Megabank Project.

Address for correspondence

Soichi Ogishima
Dept. of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University
Seiryo-cho 2-1, Aoba-ku, Sendai-shi Miyagi 980-8573 Japan
ogishima@sysmedbio.org

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1058

Non-Integrated Information and Communication Technologies in the Kidney Transplantation Process in Brazil

Alissa Peres Penteado[a], Rafael Fábio Maciel[b], João Erbs[c], Cristina Lucia Feijó Ortolani[d], Bartira Aguiar Roza[c], Ivan Torres Pisa[e]

[a]*Graduate Program in Management and Health Informatics, Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil*
[b] *Transplant Department, Instituto Social de Assistencia a Saude/Hospital Antonio Targino, Paraiba, Brazil*
[c] *Nursing Department, UNIFESP, São Paulo, Brazil*
[d]*Graduate Program in Dentistry,Universidade Paulista (UNIP), São Paulo, Brazil*
[e] *Health Informatics Department, UNIFESP, São Paulo, Brazil*

Abstract

The entire kidney transplantation process in Brazil is defined through laws, decrees, ordinances, and resolutions, but there is no defined theoretical map describing this process. From this representation it's possible to perform analysis, such as the identification of bottlenecks and information and communication technologies (ICTs) that support this process. The aim of this study was to analyze and represent the kidney transplantation workflow using business process modeling notation (BPMN) and then to identify the ICTs involved in the process. This study was conducted in eight steps, including document analysis and professional evaluation. The results include the BPMN model of the kidney transplantation process in Brazil and the identification of ICTs. We discovered that there are great delays in the process due to there being many different ICTs involved, which can cause information to be poorly integrated.

Keywords:

Kidney transplantation (MeSH E02.870.500); Systems integration (H01.770.787).

Introduction

The entire kidney transplantation process in Brazil is defined through laws, decrees, ordinances, and resolutions [1], but there is no defined theoretical map describing the kidney transplantation process. Business process modeling notation (BPMN) is the language most often used for representing processes diagrammatically [2]. It facilitates understanding among all the professionals involved in the process, and also helps organizations to measure their activities. The aim of this study was to identify the ICTs involved in the kidney transplantation process (KTP) in Brazil after the process has been analyzed and represented using BPMN.

Methods

This study was conducted in eight steps. Step 1: identification of the official digital documents relating to the KTP. Step 2: analysis of the the most important official documents (http://goo.gl/KbvHpp) using the methodology analysis on documents and processes. An initial representation of the KTP with ICTs identification was created using the Bizagi software (bizagi.com). Step 3: specialist evaluation using the Delphi methodology. Step 4: incorporation of a second official document (http://goo.gl/2a6HPi) and the assessment made by the specialist. Step 5: conduction of a survey to analyze and validate the second version of the KTP with four specialists. Step 6: research on the official website of each Brazilian state for additional legislation. Step 7: analysis of 11 documents using the methodology analysis on documents and processes. Step 8: conduction of an online survey to analyze and validate the results with the transplantation state units of all 26 Brazilian states and the federal district.

Results

The result of this work as a whole is much broader since it had to map the whole kidney transplantation process in Brazil (http://goo.gl/8VOH02). In total we analyzed 13 digital documents that resulted in two processes with 45 activities and events, six organizations involved, and six different stages. The ICTs identified included the federal IT system, phone (including mobile), fax, and email. We identified that from a total of 45 activities and events, seven are supported by a federal IT system, five by telephone, two by fax, and five by email. Thirty-three activities and events have no ICT supporting their implementation.

Conclusion

We note that there is still a great delay in the process, as there are diferent types and standards of ICTs involved, which can cause disparate data. This can result in rework, information loss, and delay in a process that could be conducted in a different and more efficient way. Adopting a single system for all entities involved in the process would be the best way to meet the process needs.

References

[1] Medina-Pestanam JO, Galante NZ, Tedesco-Silva H Jr, Harada KM, Garcia VD, Abbud-Filho M, Campos Hde H, Sabbaga E. Kidney transplantation in Brazil and its geographic disparity, J. Bras. Nefrol. 2011 Dec;33(4):472-84.

[2] Falcioni D, Polini A, Polzonetti A, Re B. Direct Verification of BPMN Processes through an Optimized Unfolding Technique. In: Proceedings of the Quality Software International Conference; 2012; IEEE; p. 179–188.

Address for correspondence

Alissa Peres Penteado, master's student. Rua Botucatu, 862. Cep: 04023-062 Edifício Leal Prado, térreo. São Paulo, SP – Brazil. Phone: +551155764347 | Mobile: +5511991451124 . Email: alissappenteado@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1059

Are we talking about the same patient?

Khawllah Roussi[1], Vanessa Soussa[2], Karen Dunn Lopez[2], Abhinaya Balasubramanian[3], Gail M. Keenan[4], Michel Burton[5], Neil Bahroos[5,6], Barbara DiEugenio[3], Andrew D. Boyd[1]

[1]Department of Biomedical and Health Information Sciences, University of Illinois at Chicago (UIC), Chicago, Illinois, USA [2]Department of Health System Science, UIC [3]Department of Computer Science, UIC [4]Department of Health Care Environments and Systems, University of Florida, Gainesville, FL, USA [5]Department of Medicine, UIC [6]Office of the Vice Chancellor of Research, UIC

Abstract

The objective of this study is to determine the degree of similarities between the clinical terms used by physicians and nurses in their documentation.

Keywords:

NLP, NANDA, NOC, NIC, Health Informatics, Physician Documentation, Interprofessional Communication

Introduction

A coherent electronic record of patient care can aid shared understanding among health care team members and can improve patient and population outcomes. This point was illustrated when, lack of physician nurse documentation integration recently made international news when an Ebola patient was sent home in Dallas due to a lack of physician-nurse documentation sharing [1]. This lack of integration is commonplace in EHRs, including discharge documents. When patients are discharged, only the physician perspective is included in discharge summary. However, nurses provide the vast majority of direct care in the hospital. Nursing care can be documented in structured terminologies: NANDA (nursing diagnosis), NIC(nursing intervenitoins) and NOC (nursing outomes) for outcomes [2]. The similarities between the two profession's individual terms and terminologies is poorly understood and have been infrequently evaluated.

Methods

We utilized a previously de-identified published list [3] of 53,423 physician concepts, created from 8 years of physician discharge summaries. These concepts were assigned concept unique identifier (CUI) terms by MedLEE (a medical information extraction program) [3]. The complete list of 1,011 terms in NANDA-I, NIC, and NOC was generated by the HANDS [2] nursing plan of care software. Using the Unified Medical Language System (UMLS), each nursing term was mapped to a corresponding CUI [3]. The physician and nursing CUIs were assessed for being synonyms via UMLS.

Results

The 1,011 nursing terms map to 956 disctinct UMLS CUI concepts. Physicians only use 21% of these 956 nursing CUI concepts which represent a mere 0.4% of all CUI terms used by physicians in their documentation. Specifically, 28% of the NANDA-I terms, 13% of the NOC terms, and 24% of the NIC terms mapped to physician terms.

Discussion

The fact that only 21% of the nursing concepts are reflected in physician documentation demonstrates true differences in the professional practices. Prior work analyzing individual patients documentation demonstrated 17.4% of synonyms between MD/RN [4]. This study differs from others in that we specifically focus on examining the overlap of documentation terms used by physicians and nurses using comprehensive vocabularies that represent each profession reflecting the larger diversity of patient conditions.

Conclusion

The large discrepancy found between terms used by doctors versus nurses suggests the alarming potential for misunderstandings between the two professions. The finding also supports the assertion that nursing has a focus that is distinct from medicine, demonstrating a clear need for nurses to document care in the patient record so that the care is interoperable (can be analyzed to evaluate impact). The benefit of similarities calculations is the future potential for improving collaborative practice by ensuring concepts that are different are understandable to both professions. One limitation is that this study was conducted in a single hospital. Additionally, we evaluated hospital-wide usage and not term frequency. However,the findings indicate that both professions must be aware of differences in use of specific terms.

References

[1] Bever L. Dallas hospital blames 'flaw' in 'workflow' for release of Ebola patient as a more complete picture of his travels emerges. *Washington Post*, October 2, 2014.

[2] Keenan GM, Yakel E, Yao Y, Xu D, Szalacha L, Tschannen D, Ford Y, Chen Y, Johnson A, Dunn Lopez K, and Wilkie DJ. Maintaining a consistent big picture: Meaningful use of a web-based POC EHR system. *Int J Nurs Knowl.* 2012: 23(3): 119-133.

[3] Burton M, Sadhu E, Dong X, Lei Z, and Bahroos N. Clinical Research Data Warehouse Enrichment Via NLP Over Discharge Summaries. *Proceedings AMIA Clinical Research Informatics*, Apr 10, 2014, San Francisco CA.

[4] So EY and Park H. Exploring the possibility of information sharing between the medical and nursing domains by mapping medical records to SNOMED CT and ICNP. *Healthc Inform Res*, 2011: 17(3): 156-161.

Address for correspondence

1919 West Taylor Street Chicago, Illinois 60612, USA
Andrew D. Boyd. boyda@uic.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1060

An HL7-FHIR-based Object Model for a Home-Centered Data Warehouse for Ambient Assisted Living Environments

Jonas Schwartze[a], Lars Jansen[b], Harald Schrom[b], Klaus-Hendrik Wolf[a], Reinhold Haux[a], Michael Marschollek[a]

[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig and Hannover Medical School, Germany*
[b] *Institute of Computer and Network Engineering, University of Braunschweig, Germany*

Abstract

Current AAL environments focus on assisting a single person with seperated technologies. There is no interoperability between sub-domains in home environments, like building energy management or housing industry services. BASIS (Building Automation by a Scalable and Intelligent System) aims to integrate all sensors and actuators into a single, efficient home bus. First step is to create a semtically enriched data warehouse object model. We choose FHIR and built an object model mainly based on the Observation, Device and Location resources with minor extensions needed by AAL-foreign sub domains. FHIR turned out to be very flexible and complete for other home related sub-domains. The object model is implemented in a separated software-partition storing all structural and procedural data of BASIS.

Keywords:

Ambient Assisted Living; HL7 FHIR; Data Warehousing.

Introduction

In an aing society, Ambient Assisted Living (AAL) has developed as a strategy to handle rising healthcare costs and sinking independence and self-reliance of elderly. Most existing AAL environments focus an assisting a single person with a composition of seperated home centered health-enabling techonolgies. [1] The BASIS project (Building Automation by a Scalable and Intelligent System) integrates sensors and actuators of all technical sub-domains in houses coplementary to AAL like building energy management, home safety, smart grid integration and housing industry. The architecture provides a central building manager with several meshed segment controller, each controlling a bus segment with any number of bus nodes (ref. [2]). Therefore, a data storage built upon a common domain model is needed. The main objective of this paper is to propose an HL7-FHIR-based resource model as foundation of a cross-domain data warehouse for home and building environments.

Methods

First, we defined a set of use cases representing each relevant sub-domain, in cooperation with the project partner. Socio-technical participants and domain related concepts were extracted to be reflected as possible object in the object model. Second, we listed existing domain-specific sensors and actuators with their measured data units or actions triggered.

Results

Since all use cases rely on saving, processing and reading (virtual) sensor data we used the resouce *Observation* as central starting point. Status and reliability are mandatory and derived from the underlying bus telegrams which deliver a validity bit to identify faulty measurements and possibly broken sensors. The element performer is fixed to a *Device* resource which in turn is used to represent a bus node. Actors in all use cases can be devided into persons (e.g. residents, sub-domain staff) or devices, which are sensors, actuators and complex devices. These devices are installed at certain locations (e.g. rooms, buildings) beeing part of certain other locations, modeled by the *Location* resource, which mainly relies on *position* and *part-of* to make locations nestable. To define more detailed technical data needed by most domains, we defined the extensions *technical*, *boundary* and *numbering*.

Conclusion

We presented an architecture and object model for a cross domain house bus system useable in AAL and other environments with a common object model as foundation of a home centered data warehouse. Whith the approach described we are able to store a broad range of sensor and actuator data as well as structural information of related sub-domains.

References

[1] Gietzelt M, von Bargen T, Kohlmann M,, et al. Home-centered Health-enabling Technologies and Regional Health Information Systems. An Integration Approach Based on International Standards. Methods of Information in Medicine. 2014;53

[2] Schrom H, Michaels T, Stein S, Ernst R. SmallCAN - A Reliable, Low-Power and Low-Cost Distributed Embedded System for Energy Efficient Building Automation. In: ENERGY 2011, The First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies from May 22, 2011 to May 27, 2011. Venice/Mestre, Italy; 2011. p. 13–18.

Address for correspondence

Jonas Schwartze, M.Sc.; Peter L. Reichertz Institute for Medical Informatics; Mühlenpfordtstraße 23, D-38106 Braunschweig;

T +49 (0)531 391-9513; Jonas.Schwartze@plri.de

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1061

Data Curation: Improving Environmental Health Data Quality

Lin Yang[a], Jiao Li[a], Li Hou[a], Qing Qian[a]

[a] *Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China*

Abstract

With the growing recognition of the influence of climate change on human health, scientists' attention to analyzing the relationship between meteorological factors and adverse health effects. However, the paucity of high quality integrated data is one of the great challenges, especially when scientific studies rely on data-intensive computing. This paper aims to design an appropriate curation process to address this problem. We present a data curation workflow that: (i) follows the guidance of DCC Curation Lifecycle Model; (ii) combines manual curation with automatic curation; (iii) and solves environmental health data curation problem. The workflow was applied to a medical knowledge service system and showed that it was capable of improving work efficiency and data quality.

Keywords:

Environmental Health Research; Data Integration; Curation Workflow.

Introduction

The impact of climate change on public health, especially the increased severity of chronic air pollution exposure in areas, triggered numerous association studies. Since research has become heavily dependent on open access data, researchers need high quality data to support analysis. Data curation, which maintains quality, adds value, and provides reuse, could provide the necessary data quality. In this study, we construct a standard environmental health data curation workflow to achieve a high quality dataset for sharing.

Methods

In typical environmental health studies, researchers need to deal with data cleaning and processing case by case. We formulated this procedure in a standardrized workflow (see Figure 1) referring to Digital Curation Center (DCC) Curation Lifecycle Model. The workflow consists of six key compononts:

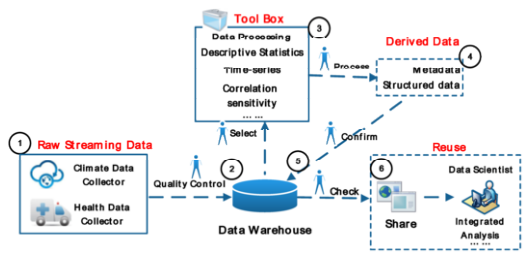


Figure 1 - The Environmental Health Data Curation Workflow

1. **Data Collection.** Raw, streaming data is collected from climate data and health data collectors.
2. **Data Ingest.** After quality assessment, curators upload qualifying data to a repository.
3. **Data Processing.** Computational tools help to clean the data and make basic analysis such as descriptive statistics, correlation analysis, time-series analysis, and sensitivity analysis. The processed data with statistical resutls is released.
4. **Data Structured Description.** Curators create metadata to describe the context of data.
5. **Data Storage.** Transform-derived data into long-term preservation formats and store for reuse.
6. **Data Reuse.** Share data via a online system for data scientists to access, integrated analyze, and promote knowledge transform.

Results

We applied the workflow to curate climate data and hosptial registry data. Climate data were collected from China ministry of environmental protection, including hourly air pollutants monitoring data and daily weather data. Registration data was from hospital emergency departments of a metropolis in China. Furthermore, the curated dataset was shared via a medical knowledge service system (http://med.ckcest.cn/) and the workflow provides backend support for the environmental health part of the system. The result shows that the workflow could help standardize environmental health data process, improve work efficiency and data quality.

Acknowledgments

This research is supported by the program of China Knowledge Center for Engineering Sciences and Technology, the National Population and Health Scientific Data Sharing Program of China, and Central Research Institutes of Basic Research and Public Service Special Operations (14R0105).

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1062

Development of a Dynamic and Adaptive Simulator for Health

Adriana L Correa-Arango[a], Carolina Tamayo-Correa[a], David Mejía-Zapata[a,b], Edison Castrillón[a], Ever A Torres-Silva[a], Ivan F Luna-Gomez[a], Natalia Restrepo[a], Sebastián Vélez-Zuluga[a], Jose F Florez-Arango[a,c,d], Jack Smith[a,e]

[a] *Escuela de Ciencias de la Salud, Universidad Pontificia Bolivariana, Colombia*
[b] *Clínica Universitaria Bolivariana, Colombia*
[c] *Facultad de Medicina, Univesidad de Antioquia, Colombia*
[d] *Hospital Pablo Tobón Uribe, Colombia*
[e] *Texas A&M University, TX, USA*

Abstract

One of the ways to develop health simulations is through the use of computers. This paper presents the use of Intelligent Computer-Aided Instruction (ICAI) for the development of an interactive simulator for learning Cardio Pulmonary Resucitation (CPR) which incorporates online tutorials, training and evaluation.

Keywords:

Intelligent Computer-Aided Instruction (ICAI); Cardio Pulmonary Resucitation (CPR); Simulation; Guidelines Computer Assited-Instruction (GCAI).

Introduction

We developed a dynamic and adaptive case simulator that complements an online CPR course. WEST [1] and WUSOR [2] were early successes using health simulators. Early ICAI research demonstrated successful engineering of such systems to determine what knowledge a student was ready to acquire, and judge effective ways to intercede in the flow of problem solving. Our dynamic and adaptive case simulator is combined with a computer program based on protocol guidance to adapt to cases in a dynamic way.

Methods

Our system is an adapted version of the general ICAI architecture with the addition of an embedded GCAI for creating patient cases. The architecture for our system is:

1. Expert Knowledge Sources (KSs): generates problems and evaluates the student's solutions and steps. A representation of correct treatment protocol is parameterized, and is used to create a difference model of the student versus expert behavior.
2. Student-Model KSs: contains data about the student's interactions, selections, and other behaviors, along with a representation of understanding of the taught material. The appropriateness of each student action in the context of the case being solved and the recorded behavior differences is evaluated.
3. Tutor KSs: represents the data, information and knowledge to integrate the students understanding, the teaching method, and the subject area.

Figure 1– Achitecture of our coaching system for life support training

Results

Our tutoring KSs are successful in implementing the coaching pegadogy in CPR simulation based training. Our empirical approach to determined the data, information and knowledge required for the Tutoring Module resulted in an adapted version of the general ICAI architecture with the addition of an embedded GCAI for creating cases using a coaching pegagogy.

Discussion

The adapted version of the general ICAI architecture is now being used for other projects and is generalizable outside the CPR simulation context. One of these additional contexts is the design of a model of service delivery for both mother and fetus mediated by information and Communication Technologies(ICT) which enables monitoring of prenatal care, timely identification and management of high-risk obstetrics, and remote assistance for complicated situations.

References

[1] Atkinson, R.C. and Wilson, H.A. (Eds.). 1969. Computer-assisted instruction. New York: Academic Press.
[2] Goldstein, I.G., 1982 The genetic graph: a representation for the evolution of procedural knowledge. In D. Sleeman and J.S. Brown (Eds.). Intelligent tutoring systems. London, Academic Press

Address for correspondence

Florez-Arango, Jose F. (jose.floreza@upb.edu.co)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1063

On Analyzing Readmissions Using A Trajectory Model: Evidence From Israel

Ofir Ben-Assuli[a], Rema Padman[b], Martha Bowman[c], Moshe Leshno[d], Itamar Shabtai[e]

[a]Ono Academic College, Israel; [b]Carnegie Mellon University, Pittsburgh, USA; [c]Olmsted Medical Center, Minnesota, USA; [d]Tel-Aviv University, Israel; [e]College of Management Academic Studies, Rishon Lezion, Israel

Abstract

The problem of readmission, wherein patients are readmitted for the same or a related condition shortly after discharge, has become a challenge worldwide from care quality and financial perspectives. In this study, we explore 30-day readmission data for predicting who is likely to be readmitted and understanding key factors contributing to preventable readmissions using the developmental trajectory of creatinine level as a key laboratory marker of serious illness and a potential predictor of future readmission. Using Electronic Health Record data on 928 patients over ten different visits to emergency departments across Israel, we apply a semi-parametric, statistical, group-based trajectory model to elicit three distinct creatinine-based trajectories over time with differing 30-day readmission rates for males and females. Analysis of readmission risk stratification of the patient population using other relevant factors is ongoing research.

Keywords:

Readmission Prediction; Group-Based Trajectory Model; Creatinine Level; Electronic Health Record

Introduction

Readmission is the process of being admitted to a hospital, usually within 30-days of a patient's initial admission, costing healthcare systems billions of dollars every year [1]. Reducing hospital readmissions has long been a health policy goal because it can lower healthcare costs, improve quality, and increase patient satisfaction. In Israel, readmission rates have been relatively stable in the last decade, but at a fairly high level of 14% [2]. This study proposes a novel approach for predicting the risk of readmission and assessing the key factors leading to readmission that may facilitate improved patient stratification and follow up procedures.

Methods

Leveraging the availability of a rich and unique clinical dataset extracted from the EHR of the largest HMO in Israel, this preliminary study analyzes readmission risk prediction using semi-parametric, statistical, group-based trajectory models (GBTM) estimated using creatinine level as a specific clinical marker of serious illness [3, 4]. In particular, we examine how readmission risk may progress, based on creatinine levels, over 10 separate hospital admissions over time for an identified patient population of 928 patients presenting in the emergency department, what the different trajectories and their frequencies are, and the relationship of these trajectories to patient characteristics.

Results

Figure 1 shows the baseline model of progression of patients' health condition. There are three distinct trajectories describing the likelihood of readmission based on creatinine levels in the blood. Group 2, the first level, in stable state, includes almost 85% of the population. Group 1, with almost 13% of the population, has progressively worsening health conditions. Group 3 has only 2.5% of patients but is the most complex, with high average levels of creatinine in the blood, which seems to level off after 8 visits. By analyzing patient and visit characteristics to better understand factors that may be driving the readmission risk, we observe that the complex Group 3 also had the lowest readmission rate by gender. This non-intuitive result needs to be investigated further to better understand the drivers of risk in each patient subgroup.

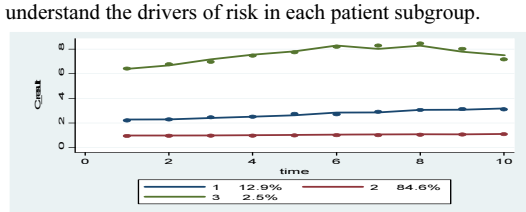


Figure 1. Baseline Model Results

Conclusion

Our early findings suggest that creatinine level based risk stratification can identify distinct groups of patients based on their presentation in the emergency department for admission and readmission purposes. Ongoing research will investigate additional time-invariant risk factors such as diagnoses and time-varying factors such as critical vitals and other laboratory values to improve risk stratification of this patient population.

References

[1] Joynt KE, Jha AK. Thirty-day readmissions — truth and consequences. N Engl J Med. 2012;366(15):1366–9.

[2] Haklai Z, Maron J, Aborva M, Gordon S, Zeltz I, Apelboim Y. Admissions to Internal Departments 2000-2012. In: The Ministry of Health I, editor. Jerusalem 2013.

[3] Nagin DS. Group based models of development. Boston, MA: Harvard University Press; 2005.

[4] Padman R, Nagin D, Xie Q. Disease Progression and Risk Prediction for Chronic Kidney Disease: Analysis of Electronic Health Record Data using Group-Based Trajectory Models. WITS 2014 Proceedings. Auckland, New Zealand.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1064

Automated Detection of Health Websites' HONcode Conformity: Can N-gram Tokenization Replace Stemming?

Célia Boyer[a], Ljiljana Dolamic[a] , Natalia Grabar[b]

[a]Health on The Net Foundation, Geneva, Switzerland, [b] Université Lille 3, France

Abstract

Authors evaluated supervised automatic classification algorithms for determination of health related web-page compliance with individual HONcode criteria of conduct using varying length character n-gram vectors to represent healthcare web page documents. The training/testing collection comprised web page fragments extracted by HONcode experts during the manual certification process. The authors compared automated classification performance of n-gram tokenization to the automated classification performance of document words and Porter-stemmed document words using a Naive Bayes classifier and DF (document frequency) dimensionality reduction metrics. The study attempted to determine whether the automated, language-independent approach might safely replace word-based classification. Using 5-grams as document features, authors also compared the baseline DF reduction function to Chi-square and Z-score dimensionality reductions. Overall study results indicate that n-gram tokenization provided a potentially viable alternative to document word stemming.

Keywords:
Machine learning; N-gram; HONcode.

Introduction

The HON Foundation's Code of Conduct[1], comprised of eight procedural guidelines, helps to indicate the credibility of online health information for both website editors and users. Currently, HON expert reviewers manually assess, re-assess, and certify health websites for compliance with the HONcode principles of conduct. The authors report herein a feasibility study to determine whether a specific machine learning algorithm based on n-gram representation of a document's content, can assist in the HONcode certification process.

Methods

In the scope of KHRESMOI[2] project we have developed a system for automated detection of HONcode principles [1]. Based on previous experience [1], the Naive Bayes machine learning algorithm is chosen in this study to compare the results obtained when various size n-grams (e.g. C3, C4, C5) and stems (W1p) are used as document tokens to the results of the word tokens (W1) baseline. The goal was to determine the extent to which words might be replaced as tokens while not sacrificing system classification performance. Keeping 30% of top ranked features, we also explored DF, Chi-square and Z-score dimensionality reduction for the 5-gram tokenization.

Results

Authors chose precision (P), recall (R) and F_1-measure to represent the quality of the classification. The results indicate that the tokenization method that produces the best precision results varies for each HONcode criterion. While W1 tokenization provides highest precision and highest F_1 value for "Authority" (0.64; 0.69) or "Privacy" (0.91; 0.94), highest precision for the "Reference" is obtained for 5-gram tokenization. For the "Justifiability" criterion the most balanced precision/recall tradeoff is achieved by C3.

The relative difference in precision between W1p and C5 spanning from -7.25% to 3.45% as well as similar behavior to W1 or W1p in relation with dimensionality reduction (e.g. features kept: 80% vs. 30%. Precision loss W1: -11.08%; W1p: -10.28%; C5: -10.41%) indicate the usability of C5 as an alternative to stem or word for the English language.

Results also show that both DF and Z-score dimensionality reduction significantly outperformed the Chi-square in terms of precision, with the exception of the "Contact details" criterion for the C5 tokenization.

Conclusion

Our study results indicate that the n-gram approach is a viable alternative to both word and stem tokenization. Choosing the "correct" dimensionality reduction algorithm can additionally improve the classification results. Accounting for importance of linguistic treatment for morphologically complex languages, and the baseline established here, one can suppose that the language-independent n-gram approach would not only be interchangeable but also might result in better performance for languages other than English.

References

[1] Boyer C, Dolamic L. Feasibility of automated detection of honcode conformity for health related websites. IJACSA 2014: Mar; 5(3):69-74

Address for correspondence

Célia Boyer, celia.boyer@healthonnet.org. Health on The Net Foundation. Chemin du Petit-Bel-Air 2, 1225 Chêne-Bourg, Switzerland
+41 (0) 22 37 26 250

[1] http://www.healthonnet.org/HONcode/Conduct.html
[2] European project KHRESMOI (2010-2014, project No. 257528), http://khresmoi.eu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1065

Data Science Solution to Event Prediction in Outsourced Clinical Trial Models

Daniel Dalevi[a], Susan Lovick[b], Helen Mann[b], Paul D. Metcalfe[a], Stuart Spencer[b], Sally Hollis[b], David Ruau[a]

[a]*Advanced Analytics Centre, Biometrics and Information Sciences, AstraZeneca, Sweden/UK*
[b]*Biometrics and Information Sciences, AstraZeneca, UK*

Abstract

Late phase clinical trials are regularly outsourced to a Contract Research Organisation (CRO) while the risk and accountability remain within the sponsor company. Many statistical tasks are delivered by the CRO and later revalidated by the sponsor. Here, we report a technological approach to standardised event prediction. We have built a dynamic web application around an R-package with the aim of delivering reliable event predictions, simplifying communication and increasing trust between the CRO and the in-house statisticians via transparency. Short learning curve, interactivity, reproducibility and data diagnostics are key here. The current implementation is motivated by time-to-event prediction in oncology. We demonstrate a clear benefit of standardisation for both parties. The tool can be used for exploration, communication, sensitivity analysis and generating standard reports. At this point we wish to present this tool and share some of the insights we have gained during the development.

Keywords:
Event prediction; oncology; visualizations; R; Shiny

Introduction

Time to event analysis is generally used when assessing the efficacy of newly developed oncology drugs. In these studies, the timing of analysis is often based on reaching a target level of events. Therefore it is desirable to predict the times of these landmark events for logistic planning – before and during the course of the trial [1]. Typical design related information are: length of trial, recruitment period, shape of recruitment, randomization balance, expected rates and statistical test parameters (alpha, power etc). In the case of an ongoing trial we also have accumulated data. We have implemented a web application using RStudio Shiny server together with an R-package simplifying the analysis, the learning curve and hand-over time while increasing interactivity, reproducibility, quality and trust.

Materials and Methods

This section describes some of the theory behind the event predictions analysis. Non-linear recruitment is used together with a survival function for describing the time-to-events. Constant or monotonically increasing or decreasing risks are modelled using exponential or Weibull distributions. There are two prediction modes. In the "from parameters" mode, where there is little or no data, study design parameters are used to specify an exponential model as the basis to predict the required number of events, the critical hazard ratio, predicted time points, the power and relative risks. In the "from data" mode, a Weibull model is fitted to the accumulated data.

Predictions for patients without an event is generated by simulating new event times. All methods are implemented in R and Shiny server is used to produce the web-interface.

Results

A simple use-case: a CRO statistician investigates an ongoing clinical trial with 979 patients currently recruited and 699 events observed. Using the online web interface the model predicts that the 900 events target will be reached in September 2016 with a confidence interval of [Aug 2016, Oct 2016]. The CRO sends the report to the study team who can easily reproduce and validate the results. The interactivity allows easy exploration of the prediction and gives an immediate feel for the sensitivity towards changes in settings.

Conclusion

Enabling CRO and Pharmaceutical companies to work on harmonised statistical computing platforms developed in-house has the potential to reduce data transfer cost and communication time. Event prediction in oncology trials can be complex and this approach enables easy diagnostics and sensitivity analysis with a low learning curve. The tool can be used for exploration, communication, sensitivity analysis and for generating standard reports, as well as facilitating single-point implementation of updates, bug fixes and expansions.

Acknowledgments

We would like to thank Andrew Stone for his support.

References

[1] Ying, G-S., Heitjan, D.F. Weibull Prediction of Event Times in Clinical Trials. Pharmaceuticals Statistics (2008) 7 107-1.

Address for correspondence

Daniel Dalevi, daniel.dalevi@astrazeneca.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1066

Improving Hospital-wide Patient Scheduling Decisions by Clinical Pathway Mining

Daniel Gartner[a], Ines V. Arnolds[b] and Stefan Nickel[b]

[a] *The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA*
[b] *Institute for Operations Research, Discrete Optimization, Karlsruhe Institute of Technology, Karlsruhe, Germany*

Abstract

Recent research has highlighted the need for solving hospital-wide patient scheduling problems. Inpatient scheduling, patient activities have to be scheduled on scarce hospital resources such that temporal relations between activities (e.g. for recovery times) are ensured. Common objectives are, among others, the minimization of the length of stay (LOS). In this paper, we consider a hospital-wide patient scheduling problem with LOS minimization based on uncertain clinical pathways. We approach the problem in three stages: First, we learn most likely clinical pathways using a sequential pattern mining approach. Second, we provide a mathematical model for patient scheduling and finally, we combine the two approaches. In an experimental study carried out using real-world data, we show that our approach outperforms baseline approaches on two metrics.

Keywords:
Artificial Intelligence; Linear Programming

Introduction

Hospital resources become more and more scarce which challenges hospital-wide coordination of patient care. To plan treatments throughout the patients' LOS, high-quality information about planned activities and time lags between them, henceforth denoted as clinical pathways (CP), is necessary. When information about patients is incomplete, effective and efficient planning of care processes is difficult.

CPs support consistent application of evidence-based medicine for the best patient outcomes. Often this has the effect of placing an emphasis on reduction of unwarranted variation in clinical practice [4]. A brief overview of similarity measures of and adherence to CPs are presented in [1]. In this poster, we present and evaluate a new and unusual way of predicting CPs. Similar to Gartner and Arnolds [1], we first learn significant CPs from large transactional data and incorporate them into a mathematical model. The novelty is, however, that we perform patient scheduling in addition to layout decisions.

Methods

Learning significant CPs from data can be seen as a sequential pattern mining task. One out of many algorithms that tackle the problem is described in [1] It also provides an overview of related work. A probabilistic prefix-tree acceptor (PDFA) is learned from data. States are merged recursively by using a similarity measure, and the diagram probability can be inferred and interpreted as the transition probability between specialties. We incorporate most likely transitions into an extension of a patient flow problem with fixed admission dates presented in [2] which can verbally be described as follows: Binary decision variables decide whether or not an activity is scheduled for a period. The objective function minimizes LOS. Precedence constraints capture clinical pathways, resource constraints for overnight resources capture scarce beds. Further constraints ensure that each activity is assigned to exactly one period of its time window.

Results

We extracted the specialty flow from data of 16,000 patients obtained from a German academic teaching hospital. Our computational study reveals that using IBM ILOG CPLEX concert technology, the computation time for each cross-validation run takes less than 1 minute on a 2.4 GHz PC. Our mean absolute error and LOS deviation reductions range between 2% and 5%, depending on the PDFA generalization.

Conclusion

In this poster, we introduce the connection between a sequential pattern mining approach and a patient scheduling problem. We show that our approach outperforms baseline approaches on two evaluation measures. Future work may incorporate our approach into the platform presented in [3].

References

[1] Gartner D and Arnolds IV. Clinical Pathway Mining for Downstream Patient Flow Planning. Proceedings of the 2014 INFORMS Workshop on data mining and analytics, San Francisco 2014.

[2] Gartner D and Kolisch R. Scheduling the Hospital-wide Flow of Elective Patients. European Journal of Operational Research 2014:223(3): 689-699.

[3] Gartner D and Padman R. E-HOSPITAL – An Environment for Hospital Operations and Services Planning using Information Technology and Algebraic Languages. MEDINFO 2015 – 15th World Congress on Health and Biomedical Informatics, 2015.

[4] Wennberg JE, Blowers L, Parker R, Gittelsohn AM. Changes in Tonsillectomy Rates Associated With Feedback and Review. Pediatrics 1977:59(6):821-826.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1067

Annotation methods to develop and evaluate an expert system based on natural language processing in electronic medical records

Quentin Gicquel[a], Nastassia Tvardik[a], Côme Bouvry[a], Ivan Kergourlay[b], André Bittar[c], Frédérique Segond[d], Stefan Darmoni[b], Marie-Hélène Metzger[a,e]

[a] *Université Lyon 1, UMR CNRS UCBL 5558, Lyon, France*
[b] *Department of Biomedical Informatics, Rouen University Hospital, TIBS, LITIS EA 4108 Rouen University, France*
[c] *Holmes Semantic Solutions, Grenoble, France*
[d] *Viseo Technologies, Grenoble, France*
[e] *Hospices Civils de Lyon, Groupement Hospitalier Nord, Lyon, France*

Abstract

The objective of the SYNODOS collaborative project was to develop a generic IT solution, combining a medical terminology server, a semantic analyser and a knowledge base. The goal of the project was to generate meaningful epidemiological data for various medical domains from the textual content of French medical records. In the context of this project, we built a care pathway oriented conceptual model and corresponding annotation method to develop and evaluate an expert system's knowledge base. The annotation method is based on a semi-automatic process, using a software application (MedIndex). This application exchanges with a cross-lingual multi-termino-ontology portal. The annotator selects the most appropriate medical code proposed for the medical concept in question by the multi-termino-ontology portal and temporally labels the medical concept according to the course of the medical event. This choice of conceptual model and annotation method aims to create a generic database of facts for the secondary use of electronic health records data.

Keywords:

Natural language processing; Knowledge base; Medical records; Information Storage and Retrieval Methods; Semantics; Terminology;

Introduction

Most information extraction technology requires a corpus of annotated medical records for development and evaluation. Our annotation approach was clinically-oriented in order to develop and evaluate medical expert rules involved at different levels of the solution. The objective of this paper is to describe how we developed a conceptual model based on the manual annotation of medical documents, which will be used as "gold standard" for the development and evaluation of the knowledge base.

Methods

The first source of the data was the research project ALADIN-DTH, in which 1600 medical textual documents were extracted and de-identified from four French university hospitals. The second source of data consists of 300 medical documents of patients diagnosed with colon cancer at a French referral center for oncology care. A health extractor (ECMT V2) was used to obtain the health concepts included in the medical reports. This extractor was developed by the CISMeF team, using their cross-lingual multi-termino-ontology portal (55 terminologies).

Results

MedIndex was an application developed by UMR UCBL CNRS 5558 for the annotation of medical concepts by semi-automatic methods (Figure 1).

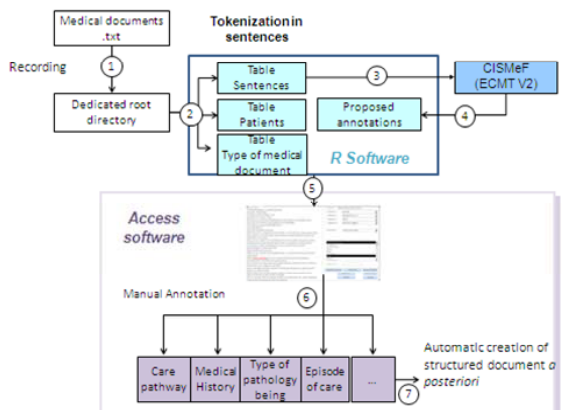


Figure 1– MedIndex dataflow

The database obtained by MedIndex is the conceptual model of the knowledge base that will be developed further for the SYNODOS solution.

Conclusion

The originality of our annotation method lies in the fact that it was not only oriented towards identifying key clinical entities mentioned in the text or determining the relationships between entities [2]. Additionally, it is oriented towards identifying medical interpretation in the context of a care pathway.

References

[1] Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I, Setzer A: Building a semantically annotated corpus of clinical texts. J Biomed Inform 2009, 42(5):950-966.

Address for correspondence

metzger.marie-helene@orange.fr

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1068

"Quartile" screening method to analyze the relationship between HIS and "AEROS" in Japan

Takayuki Hoshino[a], Nozomu Matsubara[b], Teruko Ueda[c], Yuichi Hirayama[d], Ayami Hoshino[e], Shiho Takaoka[a]

[a]*National Center for Global Health and Medicine*
[b]*Graduate School of Seigakuin University*
[c]*Aomori Public University*
[d]*System Information Partner CO., LTD*
[e]*Department of Internal Medicine, Gohyakuyama Clinic*

Abstract

Atmospheric pollution affects the health through complex mechanisms and to varying degrees. However, even by secondary usage of retrospective HIS and "AEROS" data, screening can be performed for specific ICD10 diseases caused by environmental factors. Our screening method showed good agreement with doctors' common knowledge on the relationship between atmospheric data and disease. Examples include the higher incidence of influenza at low temperatures and dry skin caused by low humidity, etc. We need to examine the method further in order to make it simpler, easier and more robust.

Keywords:
Secondary usage; HIS; "AEROS"; ICD10; "Quartile"; Screening method.

Introduction

We used two data sources for our research: AEROS (Atmospheric Environmental Regional Observation System), which is a database of government-administered real-time (by hour bat used for the daily average) regional atomospheric pollution data, including the temperature and humidity, and the in- and out-patient disease records out of HIS.

Methods

The environmental factors included in "AEROS" that were adopted for this research were the levels of 20 air pollutants (SO2, CO2, PM2.5, etc.), and the humidity and temperature for the period of 2010-2012.

The disease records comprised 520,000 cases for that period. We applied the "Quartile" screening method as follows;

1. Three quartiles are defined for each environmental item for each year, so that four quartile intervals were obtained.
2. For each quartile interval, the corresponding averages, N1 to N4, were computed for each disease in ICD10 coding.
3. Ratios Q^+=N4/N1, Q^-=N1/N4 are indicators of the intensity of the positive or negative environmental impact on each disease.
4. Data with Q^+ or Q^- greater than 1.5 during the three years were extracted.
5. 7-day moving averages were also computed to adjust for the weekday effects. Lag effects of environmental factors were adjusted for by shifting by 0 to 3 days.
6. Finally, the three-year geometric means were listed as good evidence for the correlations between environmental factors and disease. The criteria obtained thus provide us with an integrated perspective on the environment and health and the relationship between the two.

Results

Our "Quartile" screening method was considered to have high validity because of the large volume of the data (520,000 cases) and the high level of consistency of the data through the three-year period. Furthermore, our screening method showed good agreement with doctors' common knowledge on the relationship between atmospheric data and disease. For example, our results showed a relationship between

1. Influenza and low humidity (J Shamana & M Kohmn,2009),
2. Dry skin and low humidity (M Egawa,et al. 2002),
3. Ophthalmological inflammatory disease and atmospheric factors (NO2,NOX),
4. Chronic disease (cancer) and atmospheric factors (NOX) in terms of the patients' hospital visits.

Conclusion

Our "Quartile" screening method has the following four advantages in the analysis of large data: 1) simplicity, 2) readiness, 3) cost & time efficiency, and 4) robustness.

Address for correspondence

Dr. Takayuki Hoshino (e-mail address: hoshinanatsu@outlook.jp)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1069

Impact of data quality assessment on development of clinical predictive models

Jitendra Jonnagaddala [a,b,c], Siaw-Teng Liaw [a] and Pradeep Ray [d]

[a] *School of Public Health and Community Medicine, UNSW Australia*
[b] *Asia-Pacific Ubiquitous Healthcare Research Centre, UNSW Australia*
[c] *Prince of Wales Clinical School, UNSW Australia.*
[d] *Asia-Pacific Ubiquitous Healthcare Research Centre, UNSW Australia.*

Abstract

Data quality plays a very important role in predicting clinical outcomes. Data quality is multi dimensional and most relevant studies consider just one or two dimensions. In this study a systematic data quality assessment is performed using four data dimensions. The results demonstrate that performance of predictive models improves when the quality of the data is assessed and addressed systematically.

Keywords

predictive modeling; data quality assessment; data preprocessing

Introduction

Low data quality can be a serious issue in predictive modeling. The typical workflow of predictive model development includes data preprocessing, feature selection, model development and evaluation steps. There are various studies about data preprocessing and the effects of techniques employed on overall prediction problem [1, 2, 3]. However, these studies did not preprocess data from a systematic data quality assessment (DQA) perspective. Thus, the objectives of this study are i) to systematically assess data quality using four different data quality dimensions ii) develop predictive models using four algorithms to predict in-hospital mortality. The algorithms used are Naïve bayes (NB), Random forest (RF), Support Vector machines (SVM) and Multi-layer perceptron (MLP) and iii) present the effects of the systematic DQA on the performance of predictive models.

Methods

The authors used the dataset which was provided as part of the 2012 Physionet Computing in Cardiology (CinC2012) challenge [4]. Four different predictive models were developed to predict whether an ICU patient survives hospitalization. The selected data quality dimensions include completeness, consistency, correctness and contextual accuracy. Data quality metrics were calculated in percentages on each column of the dataset using the selected data quality dimensions. The calculated metrics were used to systematically profile the dataset to select relevant techniques (imputation or deletion or classification) to address the data quality issues [5]. All available variables in the dataset were used as features. The performance of the developed predictive models were measured using accuracy (represented in percentage), positive prediction value (PPV) and the F-score.

Results

DQA was performed on both training and test sets using completeness, consistency, correctness and contextual accuracy dimensions. DQA Results show that a significant amount of variables had more than 50% of missing data. The final results suggest that performing DQA systematically improves the performance of predictive models. The results are consistent with results reported in studies where just one or two data quality dimensions are used [3]. The RF based predictive model was the one which was most affected by the DQA and it showed marked improvement when all the data quality issues identified using data quality dimensions were rectified.

Conclusion

In this study, the authors explored the impacts of data quality on clinical predictive modelling by performing a systematic DQA. The authors observed that NB based model performance remained consistent but in the end the RF based model outperformed the rest of the models after DQA. The results also demonstrate that performance of predictive models improve when the quality of the data is assessed and addressed systematically.

References

[1] Bellazzi, R. and B. Zupan, *Predictive data mining in clinical medicine: current issues and guidelines.* International journal of medical informatics, 2008. 77(2): p. 81-97.

[2] Lin, J.-H. and P.J. Haug. *Data preparation framework for preprocessing clinical data in data mining*. in *AMIA Annual Symposium Proceedings*. 2006.

[3] Johnson, A.E., A.A. Kramer, and G.D. Clifford, *Data preprocessing and mortality prediction: the Physionet/CinC 2012 challenge revisited.* 2014.

[4] Silva, I., et al., *Predicting In-Hospital Mortality of ICU Patients: The PhysioNet/Computing in Cardiology Challenge 2012.* Computing in cardiology, 2012. 39: p. 245.

[5] Cismondi, F., et al., *Missing data in medical databases: Impute, delete or classify?* Artificial intelligence in medicine, 2013. 58(1): p. 63-72.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1070

Evaluating Methods for Identifying Cancer in Free-Text Pathology Reports Using Various Machine Learning and Data Preprocessing Approaches

Suranga Nath Kasthurirathne, BEng[a], Brian E. Dixon, MPA, PhD[b,c], Shaun J. Grannis, MD, MS[b,d]

[a]*Indiana University School of Informatics and Computing, Indianapolis, IN, USA*
[b]*Regenstrief Institute, Indianapolis, IN, USA*
[c]*Indiana University Fairbanks School of Public Health, Indianapolis, IN, USA*
[d]*Indiana University School of Medicine, Indianapolis, IN, USA*

Abstract

Automated detection methods can address delays and incompleteness in cancer case reporting. Existing automated efforts are largely dependent on complex dictionaries and coded data. Using a gold standard of manually reviewed pathology reports, we evaluated the performance of alternative input formats and decision models on a convenience sample of free-text pathology reports. Results showed that the input format significantly impacted performance, and specific algorithms yielded better results for presicion, recall and accuracy. We conclude that our approach is sufficiently accurate for practical purposes and represents a generalized process.

Keywords:

Public health reporting; decision models; ontologies; cancer; pathology; data preprocessing.

Introduction

Cancer case reporting is often delayed and incomplete [1]. Automated methods for identifying public health reportable cases can address this issue [2], yet a substantial amount of cancer case-related data are captured as free-text making it challenging to interpret [3]. We sought to assess approaches to identify cancer cases from free-text pathology reports to (a) determine whether we could achieve acceptable accuracy using a generalizable approach that does not require complex dictionaries, grammars or ontologies; (b) compare various candidate decision models; and (c) evaluate how data input format affects decision model accuracy.

Methods

We identified seven keywords associated with the presence of cancer in pathology reports. Each free text report was parsed and separate counts tabulated for the presence of each keyword either in positive or negated contexts using the Negex algorithm. We evaluated two preprocessed data input vectors. The first input vector ("raw count") contained positive counts (C_p) and negated counts (C_n) for keywords in each report. The second ("four-state") reduced these to a single value per keyword: 1=($C_p > C_n$); 2=($C_n > C_p$); 3=($C_p = C_n$) and 4=keyword absent. We evaluated logistic regression, naïve bayes (NB), k-nearest neighbor, random forest (RF), and J48 decision tree decision models implemented in Weka software version 3.6.10. The precision, recall, and accuracy was calculated for each input format combination.

Results

Each decision model and input format combination yielded satisfactory results. However, the "raw count" input format outperformed the "four-state" input for all three performance measures. The NB decision model produced statistically significant lower results for accuracy ($p < 0.01$); the remaining methods showed no difference as a group. For recall, all decision models showed no difference as a group. For precision, both RF and NB showed lower values ($p < 0.01$); the remaining methods were indistinguishable.

Discussion

Overall results indicated that the "raw count" input format outperformed the "four-state" format. Although we achieved reasonable performance while avoiding the use of complex dictionaries or ontologies, this approach occasionally failed to identify cases when text reports contained only disease specific terms and the seven generic keywords were absent. We conclude that our approach represents a generalized process that can be adapted for many additional clinical use cases, and is accurate enough for practical purposes.

References

[1] Zanetti R, Schmidtmann I, Sacchetto L, Binder-Foucard F, Bordoni A, Coza D, et al. Completeness and timeliness: Cancer registries could/should improve their performance. European journal of cancer. 2014.

[2] Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. American journal of public health. 2008;98(2):344.

[3] Fidahussein M, Friedlin J, Grannis S. Practical challenges in the secondary use of real-world data: the notifiable condition detector. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2011;2011:402-8.

Address for correspondence

S.N. Kasthurirathne, Indiana University School of Informatics and Computing, 535 W. Michigan Street, IT 475, Indianapolis, IN 46202, USA, (317)-278-4636, snkasthu@iupui.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1071

A Metadata based Knowledge Discovery Methodology for Seeding Translational Research

Cartik R Kothari[a], Philip RO Payne[a,b]

[a]*Department of Biomedical Informatics, School of Medicine, The Ohio State University, Columbus, OH, USA*
[b]*Data Analytics Collaborative, The Ohio State University, Columbus, OH, USA*

Abstract

In this paper, we present a semantic, metadata based knowledge discovery methodology for identifying teams of researchers from diverse backgrounds who can collaborate on interdisciplinary research projects: projects in areas that have been identified as high-impact areas at The Ohio State University. This methodology involves the semantic annotation of keywords and the postulation of semantic metrics to improve the efficiency of the path exploration algorithm as well as to rank the results. Results indicate that our methodology can discover groups of experts from diverse areas who can collaborate on translational research projects

Keywords:
Knowledge Discovery; Translational Research; Semantics.

Introduction

Translational science research [1] involves the collaboration of scientists from various disciplines and the identification of viable interdisciplinary groups of scientists is crucial to the success of translational research initiatives. We describe a methodology to identify interdisciplinary teams of researchers to collaborate on translational research projects as part of the Discovery Themes initiative at The Ohio State University.

Materials and Methods

Raw data for identifying cross-disciplinary research teams was gathered by compiling large lists of research documents from the last 5 years, which included: (a) funded research grant proposals and (b) abstracts of every article in every journal indexed by SCOPUS, restricting to only the top 10% of the most productive researchers in each discipline determined by Academic Analytics. From this corpus, text mining techniques compiled a set of approximately 60K keywords. Then, each keyword was categorized under one of 17 research disciplines or subheadings following the classication scheme used by the United States Library of Congress. We defined semantic metrics for each keyword taking into account its: a) specificity and b) usage frequency. Given the subheading under which a keyword was classified, specificity is the length of the path from the keyword to the root of the subject heading hierarchy. The usage count of the keyword was part of the input to the algorithm. Given the widely varying usage counts, we used a logarithm of the frequency. The semantic metric of the keyword is defined as the ratio of the specificity to frequency; it varies directly with specificity but inversely with usage frequency. The semantic distance between two keywords is the sum of the depths of the keywords and the semantic distance between the subject headings they belong to, except if the keywords belong to the same hierarchy when the semantic distance between them is zero. The path finding algorithm was then used to link researchers connected by 5 keywords in the graph using the semantic metrics to prune the search space.

Results and Conclusions

Starting with ~1400 researchers who were affiliated with The Ohio State University, paths with 5 keywords were generated to link them with ~14000 outside researchers (Figure 1).

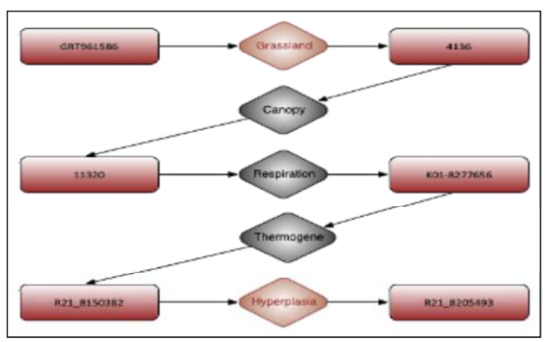


Figure 1: An example of a generated path

Our path finding algorithm is different from classical clustering techniques in that it groups data points (researchers) on the basis of their differences instead of on the basis of their similarities. A limitation of our work is the reliance on simple keywords to connect the researchers. The classification of keywords to specific subheadings and the addition of semantic metrics addresses this limitation. However, the classifications themselves are subjective; derived from manual annotation. The identity of the associated researchers could be very useful in identifying the correct context of context-sensitive keywords, given all the researchers have been de-identified in the input data set.

Acknowledgments

This research was supported by the National Cancer Institute of the National Institutes of Health under Award Number P01CA081534. We would like to acknowledge preliminary analysis work by Jason Sullivan and Julia Carpenter-Hubin.

References

[1] Woolf SH. The meaning of translational research and why it matters. JAMA 2008: 299(2): 211-213.

Address for correspondence

Cartik R Kothari Saravanamuthu, Center for Biomedical Informatics, Harvard Medical School, Longwood Campus, Countway Library, 10 Shattuck Street, Boston MA 02115. Cartik1.0@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1072

Restricted Versus Unrestricted Search Space: Experience from Mining a Large Japanese Database

Hendrik Nienhoff[a], Ursula Huebner[a], Andreas Frey[b], Mareike Przysucha[a], Michio Kimura[c]

[a] *Health Informatics Research Group, Osnabrueck University of Applied Sciences, Germany*
[b] *Nuertingen-Geislingen University, Germany*
[c] *Hamamatsu University School of Medicine, Japan*

Abstract

The aim of this study was to investigate whether standard Big Data mining methods lead to clinically useful results. An association analysis was performed using the apriori algorithm to discover associations among co-morbidities of diabetes patients. Selected data were further analyzed by using k-means clustering with age, long-term blood sugar and cholesterol values. The association analysis led to a multitude of trivial rules. Cluster analysis detected clusters of well and badly managed diabetes patients both belonging to different age groups. The study suggests the usage of cluster analysis on a restricted space to come to meaningful results.

Keywords:

diabetes mellitus; data mining; association analysis; cluster analysis.

Introduction

Data mining methods have become the core of discovering knowledge in large datasets (Big Data) [1] and have been applied in medicine [2] [3]. The aim of this study was to evaluate the usefulness of different standard data mining methods for finding useful patterns in a dataset of diabetes patients.

Methods

Based on a clinical data warehouse at the Hamamatsu University hospital, 26,890 patients were analyzed to extract patients diagnosed with diabetes type 2 (ICD10-Code 'E11'). After extensive data transformation and cleansing, association and cluster analyses were performed (using the WEKA Tool) to detect assocations of co-morbidities within the group of diabetes patients and to investigate if diabetes patients could be clustered by age, blood sugar and cholesterol values into different sub-groups. Both analyses belong to the unsupervised and describing methods of data mining [1]. Whereas the association analysis was performed on a search space of 73 variables, the cluster analysis was restricted to three variables.

Results

Using similar minimum support and confidence values as [2], 45,314 association rules were identified out of which the five strongest are shown in Table 1, ranked by confidence. Clustering with k=4 clusters showed that high levels of blood sugar went along with high cholesterol levels, each for younger and for elderly patients (Tab. 2).

Table 1– Five strongest association rules identified (n=1,339)

Rule (X→Y)	n	Supp.(X)	Conf. X→Y
H26 → H52	311	0.23	0.95
I20 → I10	320	0.24	0.84
I50 → I10	458	0.34	0.74
I50 → E78	458	0.34	0.68
E78 → I10	735	0.55	0.66

E78: Hypercholesterolaemia; H26: Other cataract; H52: Disorders of refraction; I10: Essential (primary) hypertension; I20: Angina pectoris; I50: Heart failure

Table 2– 4-Cluster centres for low-density lipoprotein

		Age		HbA_{1c}		LDL	
Clstr	n	mean	σ	mean	σ	mean	σ
1	135	68.7	7.5	8.6	1.2	85.4	25.8
2	273	75.1	5.2	6.6	0.7	60.4	17.4
3	191	56.2	8.1	6.3	0.7	66.6	19.0
4	63	39.9	9.3	8.6	1.8	89.2	26.5
total	**662**	**65.0**	**13.3**	**7.1**	**1.4**	**70.1**	**23.6**

Discussion

The association rules identified are similar to the results of Kim et.al. [2], who also detected primarily trivial rules. The cluster analysis showed that patients could be separated into two groups of diabetes patients with well or badly managed longterm blood sugar values. Well managed patients also had lower LDL values. These findings were independent of age. In summary, restricting the search space by knowledge yielded more useful clinical patterns. Data analysis was only possible after elaborate and time-consuming preprocessing.

References

[1] Maimon O and Rokach L. Data mining and knowledge discovery handbook. New York: Springer, 2005.

[2] Kim H, Shin A, Kim A and Kim Y. Comorbidity Study on type 2 diabetes mellitus using data mining. Korean Journal of internal medicine 2011; 27(2): 197-202.

[3] Valent F, Tillati S and Zanier L. Prevalence and comorbidities of known diabetes in northeastern Italy. Journal of diabetes investigation, 2013; 4(4): 355-360.

Address for correspondence

Hendrik Nienhoff, Rheiner Landstr. 35, 49078 Osnabrueck, Germany
Tel: +491786885188 E-mail: hendrik@nienhoff.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1073

Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation

Yasunobu Nohara[a], Yoshifumi Wakata[a], and Naoki Nakashima[a]

[a] *Kyushu University Hospital, Fukuoka, Japan*

Abstract

Recently, machine-learning techniques have spread many fields. However, machine-learning is still not popular in medical research field due to difficulty of interpreting. In this paper, we introduce a method of interpreting medical information using machine learning technique. The method gave new explanation of partial dependence plot and individual conditional expectation plot from medical research field.

Keywords:

Machine learning; Partial dependence plot; Individual conditional expectation.

Introduction

Recently, machine-learning technique, such as deep learning and random forest, has been spreading many fields. However, machine-learning is still not popular in medical research field. One of the reasons is difficulty of interpretation. In machine learning, prediction accuracy is a main target of the research and easy for interpretation is not sufficiently considered. In this paper, we introduce a method of interpreting medical information using machine learning technique.

Methods

Given explanatory variables $x_1, ..., x_p$, and patients' outcome y. Let N be the number of subjects. We assume predictor f is given as a black box but calculates the outcome y with high accuracy using explanatory variables $x_1, ..., x_p$ using machine learning technique. How do we study the effect of an explanatory variable x_i on patients' outcome? A naive method is comparing each explanatory variable x_i and outcome y. However, other explanatory variables also affect the outcome; therefore, we don't get accurate effect of x_i. The best way to study the effect of x_i is using the same conditions expected for the explanatory variable and investigating the difference among the outcome. It is almost impossible to recruit patients who have same age, sex, blood pressure and other conditions in a real experiment. On the other hand, we can easily calculate outcomes under various conditions on the predictor, and we use this predicted value instead of experiment result.

If we study the effect of a variable x_i, it is natural to use mean outcome under the various conditions. Let x_i^j be an explanatory variables x_i of subject-j and mean outcome fixing x_i to X is given as:

$$F_i(X) = \frac{1}{N}\sum_{j=1}^{N} f(x_1^j, x_2^j, ..., x_{i-1}^j, X, x_{i+1}^j, ..., x_p^j) \quad (1)$$

Figure 1 shows an example. If we want to compare the effect of operation-A and operation-B, we directly compare ***F(A)*** and ***F(B),*** because this comparison is simulation but under the same condition excepted for operation method.

In fact, Equation (1) is the same as the definition of partial dependence plot (PDP), which is a visualization method of high-dimensional function proposed by Friedman [1]. Individual conditional expectation (ICE) plot is an extension of PDP proposed by Goldstein [2]. ICE plot can express interaction of explanatory variables. Centered ICE (c-ICE) [2] is given as following equation (2)

$$F_i^{center}(X, x^*) = F_i(X) - F_i(x^*) \quad (2)$$

If we suppose that f is additivity, that is

$$f(x_i, \{x_j\}_{j\neq i}) = g(x_i) + h(\{x_j\}_{j\neq i}) \quad (3)$$

In this case, c-ICE is given as follows

$$F_i^{center}(X, x^*) = \frac{1}{N}\sum_{j=1}^{N}\{g(X) - g(x^*)\} \quad (4)$$

Equation (4) means that c-ICE expresses the effect of expalantory variable x_i only and does not depend on other explanatory variables if f is additivity.

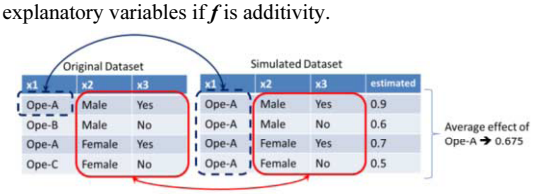


Figure 1 – Original dataset and simulated dateset of partial dependence plot

Conclusion

In this paper, we proposed a new explanation of partial dependence plot and individual conditional expectation plot from medical research field.

References

[1] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 2001: Oct 29 (5); pp. 1189-232.

[2] Goldstein A, Kapelner A, Bleich J, and Pitkin E. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. Journal of Computational and Graphical Statistics 2014. DOI: 10.1080/10618600.2014.907095.

Address for correspondence

Yasunobu Nohara, PhD

Medical Information Center, Kyushu University Hospital, Japan

3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

E-mail; y-nohara@info.med.kyushu-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1074

A Data Mining Approach to Identify Sexuality Patterns in a Brazilian University Population

Priscyla Waleska Simões[a,b], Samuel Cesconetto[a], Larissa Letieli Toniazzo de Abreu[a], Merisandra Côrtes de Mattos Garcia[a], José Márcio Cassettari Junior[a], Eros Comunello[c], Luciane Bisognin Ceretta[b], Sandra Aparecida Manenti[a]

[a] Research Group in Information and Communications Technology in Health, UNESC, Criciúma, SC, Brazil
[b] Postgraduate Program in Public Health, UNESC, Criciúma, SC, Brazil
[c] Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

Abstract

This paper presents the profile and experience of sexuality generated from a data mining classification task. We used a database about sexuality and gender violence performed on a university population in southern Brazil. The data mining task identified two relationships between the variables, which enabled the distinction of subgroups that better detail the profile and experience of sexuality. The identification of the relationships between the variables define behavioral models and factors of risk that will help define the algorithms being implemented in the data mining classification task.

Keywords:

Medical Informatics; Artificial Intelligence; Data Mining; Sexuality.

Introduction

Data mining consists of applying algorithms to identify and analyze information in order to produce patterns or models of the data [1]. The main topic of this study is sexuality, which is considered inherent to life and health, and is expressed early in humans as one of the indices that measure the level of quality of life [2]. This paper presents the profile and experience of sexuality in a university population in southern Brazil generated from a data mining classification task.

Methods

The database used for this study addressed the sexuality profile and vulnerability to sexually transmitted diseases (STDs) and acquired immune deficiency syndrome (AIDS) in a university population in southern Brazil. First, the data was pre-processed, and then the application for fuzzy clustering under development at UNESC [**3**] was used. The classification of data was carried out using the J48 algorithm in the Waikato Environment for Knowledge Analysis software (WEKA). In addition to the complete base (classification 1), an analysis of the database was examined using only males (classification 2), as well as examining a data set containing both genders but eliminating the variables of the all-male responses (classification 3). Only the rules belonging in trees with at least 70% of the records correctly classified and associations in 5% of the sample were used. We used accuracy for the evaluation of the knowledge generated.

Results

We conducted 204 experiments that resulted in 8728 rules. Classification 1 returned 3727 rules(0.76 accuracy [IC 95%=(0.73;0.80)]); classification 2 returned 1699 rules(0.75 accuracy [IC 95%=(0.71;0.79)]); and classification 3 returned 3302 rules (0.76 accuracy [IC 95%=(0.72;0.80)]). The profile obtained from the analysis was similar to that found in the literature. By observing the age of sexual activity onset, the J48 algorithm not only isolated information (15-19 years of age), but also found that this group was mostly Catholic, and the partner at first intercourse was also 15- 19 years of age.

Conclusion

Our study improved on the profile and experience of sexuality of a university population in southern Brazil by including a distinction between subgroups that allowed for more detail. This distinction may assist in the definition of the actions to be implemented for the profile and experience of sexuality of this population.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and Universidade do Extremo Sul Catarinense.

References

[1] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008;77(2):81-97.

[2] Heilborn ML. Entre as tramas da sexualidade brasileira. Rev Est Fem 2006: 14(1): 43-59.

[3] Azevedo Simões PW, Martins PJ, Casagrandre RA, Madeira K, de Mattos MC, Manenti SA et al. Using a model of parallel distributed processing associated with data mining in the characterization of sexuality in a university population. Stud Health Technol Inform. 2013;192:1135.

Address for correspondence

Prof PhD Priscyla Waleska Simões - pri@unesc.net

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1075

A Comparative Study of Bayes Net, Naive Bayes and Averaged One-Dependence Estimators for Osteoporosis Analysis

Priscyla Waleska Simões[a,b], Leandro Luiz Mazzuchello[a], Larissa Letieli Toniazzo de Abreu[a], Diego Garcia[a,c], Maitê Gabriel dos Passos[a], Ramon Venson[a], Luciane Bisognin Ceretta[b], Ana Carolina Veiga Silva[a], Maria Inês da Rosa[b], Paulo João Martins[c]

[a] *Research Group in Information and Communications Technology in Health, UNESC, Criciúma, SC, Brazil*
[b] *Postgraduate Program in Public Health, UNESC, Criciúma, SC, Brazil*
[c] *Postgraduate Program in Health Technology, Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brazil*

Abstract

This paper presents an evaluation of the accuracy of the Bayesian classifiers: Bayes Net, Naive Bayes and Averaged One-Dependence Estimator, to support diagnoses of osteopenia and osteoporosis. All classifiers showed good results, thus, given data, it is possible to produce a reasonably accurate estimate of the diagnosis.

Keywords:

Medical Informatics; Artificial Intelligence; Data Mining; Bayes Theorem; Osteoporosis.

Introduction

Osteoporosis is a disease that results in the decline in the health of the world population and challenges the public health and medical professions [1]. To address the uncertainty inherent in biomedical diagnoses, Bayesian classifiers, which are based on statistical models, have the advantage of determining the class to which a given record belongs compared to traditional classifiers, taking as base the probability of an element belonging to a class [2]. The Averaged One-Dependence Estimator (AODE) is considered an improvement on the Naive Bayes (NB) and an interesting alternative to other semi-naive approaches. It provides a good trade-off between efficiency and performance. Specifically, the AODE can be considered as an ensemble of SuperParent One-Dependence Estimators (SPODEs) because every attribute depends on the class and another shared attribute, which is designated as the superparent [3]. Thus, the diagnosis of diseases such as osteoporosis and osteopenia can become faster and precise. This paper presents an evaluation of the accuracy of the Bayesian classifiers Bayes Net (BN), NB and AODE to support the diagnosis of osteopenia and osteoporosis.

Methods

The database used in the experiments originated from a cross-sectional study involving 1871 women in southern Santa Catarina who were subjected to bone densitometry measurements under a specialized service from 2010 to 2012. This study was approved by the Ethics Committee under protocol 82939/2012. The experiments included the following steps: data cleaning and binarization; transformation of variables; selection tool; data mining; variable selection and use of classifiers. For validation was used a test set. Higher incidence of osteoporosis was presented among the 1871 women included in the sample.

Results

Approximately 300 rules were generated and selected for evaluation. Our results showed higher mean instances correctly classified using the Naive Bayes algorithm (82.84±14.42), and the average of incorrectly classified instances was higher for Bayes Net (17.46±14.76). The result for the kappa coefficient, related to the agreement between the result of the classifier and the result presented in the database by the specialist, was higher for the Bayes Net algorithm (0.33±0.29).

Conclusion

The development of this research includes the understanding that Bayesian classifiers use Bayes theorem, which allows them to stand out because of their efficiency, speed and accuracy, thus explaining its usefulness in the diagnosis process. Based on the results of analyzing statistical measures in the experiments described here (instances classified correctly and incorrectly, the kappa coefficient, mean absolute error, sensitivity, specificity, accuracy, recall, F-measure, and AUC), all classifiers showed good results, thus, given these data, it is possible to produce a reasonably accurate estimate of the diagnosis.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and Universidade do Extremo Sul Catarinense (UNESC).

References

[1] Kanis JA, McCloskey EV, Johansson H, Cooper C, Rizzoli R, Reginster J-Y. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 2012: 24 :23–57.

[2] Mello, LC. Descoberta de conhecimento em banco de dados (mineração de dados) e redes bayesianas: estado da arte, 2001.

[3] Htike ZZ, Win LS. Classification of eukaryotic splice-junction genetic sequences using Averaged One-Dependence Estimators with subsumption resolution. Procedia Comput Sci 2013: 23:36-43.

Address for correspondence

Prof PhD Priscyla Waleska Simões – pri@unesc.net

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1076

Trivalent Influenza Vaccine Adverse Event Analysis Based On MedDRA System Organ Classes Using VAERS Data

Cui Tao[a], Jingcheng Du[a], Yi Cai[b], Yong Chen[b]

[a] *School of Biomedical Informatics,* [b] *School of Public Health*
The University of Texas Health Science Center at Houston, TX, USA

Abstract

We studied serious reports following influnza vaccine from VAERS database in year 2011. Our statistical analyses revealed differences of reactions among different age groups and between genders. The results may lead to additional studies to uncover factors contributing to the individual differences in susceptibility to influenza infection.

Keywords:

Influnza vaccine adverse event, MedDRA, VAERS

Introduction

Seasonal influenza is a common vaccine-preventable disease with substantial morbidity. The social burden of seasonal flu can be substantial [1]. Annual vaccination is the most effective strategy to prevent influenza [2]. While their benefits far overweigh their risks and costs, influenza vaccines are accompanied with specific adverse events. Post-approval surveillance of vaccine adverse events is critically needed to assess the vaccine safety throughout its life on the market. The Vaccine Adverse Event Reporting System (VAERS) is a passive surveillance system to monitor vaccine safety after the administration of vaccines licensed in the United States [3]. In this study, we explored statistical analysis on annotated symptoms in the VAERS reports for patients with different genders and ages.

Materials and Methods

We searched the VAERS for US reports after Trivalent Influenza Vaccine (FLU3) in year 2011 and extracted serious reports (i.e., death, life-threatening illness, hospitalization, prolonged hospitalization, or permanent disability). For each report, the VAERS provides annotations for post-vaccination symptoms in Medical Dictionary for Regulatory Activities (MedDRA) terms [4]. To faciliate further statistical analysis, we further grouped these symptoms based on the MedDRA System Organ Class (SOC) using the NCBO Web Services [5].

To model the total number of serious symptoms per subject, we fit a zero-truncated Poisson regression on age groups and gender because a subject has to have at least one symptom to be included in the VAERS database. To study the risks of having serious symptom for each SOC type, we use logistic regression on age groups and gender.

Results

During the study period, VAERS received 7986 FLU3 reports, 638 were serious. Out of the 638 reports, 324 were for female patients, 295 were for male patients, and 134 , 156, 110, 185 were for patients in age groups 0.5-17, 17-49, 49-64, and >64 repectively. 5407 symptoms were grouped into 26 SOCs. The most frequent SOCs in the 638 reports are nervous system disorders, general disorders, and administration site condition and investigations.

Analysis using zero-truncated Poisson model indicated that the avarage number of symptoms per subjects in the study cohort is 8.74 (95% CI 6.76, 10.73). There are statistically significant differences in number of symptoms among four age groups and between different genders. The youngest age group (0.5 - 17 years) has the smallest number of symptoms per year, followed by age group 2 (17-49), age group 4 (>64), and finally age group 3 (49 – 64). The average number of symptoms for subjects of 17-49 years old is 13% higher than the average number of symptoms for subjects of 0.5-17 years old with the same gender (p=0.003).

Conclusion

This poster reports our preliminary analyses on influenza vaccine adverse events using VAERS data and MedDRA SOCs, which revealed differences of reactions among different age groups and between genders. The results may lead to additional studies to uncover factors contributing to the individual differences in susceptibility to influenza infection.

Acknowledgements

This research is partially supported by the National Library Of Medicine of the National Institutes of Health under Award Number R01LM011829 to Dr. Cui Tao.

References

1. Poehling KA, Edwards KM, Griffin MR, Szilagyi PG, Staat MA, Iwane MK, et al. The burden of influenza in young children, 2004-2009. Pediatrics. 2013; **131**(2): 207-16.
2. CDC: Children, the Flu, and the Flu Vaccine. [cited; Available from: http://www.cdc.gov/flu/protect/children.htm
3. The U.S. FDA/CDC Vaccine Adverse Event Reporting System (VAERS) [cited; Available from: http://vaers.hhs.gov/index
4. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug safety. 1999; **20**(2): 109-17.
5. Whetzel PL. NCBO Technology: Powering semantically aware applications. Journal of biomedical semantics. 2013; **4 Suppl 1**: S8.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1077

Temporal Relation Extraction in Outcome Variances of Clinical Pathways

Takanori Yamashita[a, b], Yoshifumi Wakata[a], Satoshi Hamai[c], Yasuharu Nakashima[c], Yukihide Iwamoto[c], Brendan Franagan[b], Naoki Nakashima[a], Sachio Hirokawa[d]

[a] *Medical Information Center, Kyushu University Hospital, Fukuoka, Japan*
[b] *Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan*
[c] *Department of Orthopaedic Surgery, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan*
[d] *Research Institute for Information Technology, Kyushu University, Fukuoka, Japan*

Abstract

Recently the clinical pathway has progressed with digitalization and the analysis of activity. There are many previous studies on the clinical pathway but not many feed directly into medical practice. We constructed a mind map system that applies the spanning tree. This system can visualize temporal relations in outcome variances, and indicate outcomes that affect long-term hospitalization.

Keywords:
Clinical pathway; Mind map; Hip replacement arthroplasty; Spanning Tree.

Introduction

Recently the clinical pathway has progressed with digitalization and the analysis of activity. There are many previous studies on the clinical pathway but not many feed directly into medical practice. The evaluation of temporal similarity among workflows has been reported, but they could not evaluate the medical process [1]. In this paper we visualize the temporal relations in outcome variances of the hip replacement arthroplasty clinical pathway and propose factors that affect long-term hospitalization.

Methods

A search engine was constructed from the outcomes of patients that were evaluated from admission to discharge. The patient's postoperative hospitalization days (POD) was also added as an attribute. If the POD search condition $\boldsymbol{q}$ is given, it will search for a list of patients with a POD longer than $\boldsymbol{q}$. In [2], the relationship of factors was visualized through combining generality and similarity by applying the spanning tree algorithm. We constructed a mind map system based on this method that can visualize the relationship of common attributes.

First, all of the outcomes that appear in the search results are sorted in lexicographic order of similarity and POD, and the top $\boldsymbol{K}$ outcomes $(\boldsymbol{o})$ with days $(\boldsymbol{d})$ $\{\boldsymbol{od_1}, \boldsymbol{od_2}, ..., \boldsymbol{od_K}\}$ are extracted. Second, the similarity of $\boldsymbol{o}$ and $\boldsymbol{q}$ represented by $\boldsymbol{sim(q, o)}$ was evaluated by the Jaccard index. Further, $\boldsymbol{sim(q, o_1) < sim(q, o_2)}$ is defined as (1) or (2).

$$sim(q, o_1) < sim(q, o_2) \quad (1)$$

$$sim(q, o_1) = sim(q, o_2) \quad \wedge \quad d_1 < d_2 \quad (2)$$

Then the tree will expand in the order of $\boldsymbol{od_2}, \boldsymbol{od_3}, \ldots,$ with $\boldsymbol{od_1}$ as the root of the tree. When a tree that contains the outcomes to $\boldsymbol{od_{(i-1)}}$ is obtained, the node $\boldsymbol{od_{(i)}}$ is connected to the node with the greatest similarity in $\boldsymbol{od_1}, \ldots, \boldsymbol{od_{(i-1)}}$ (Figure 1).

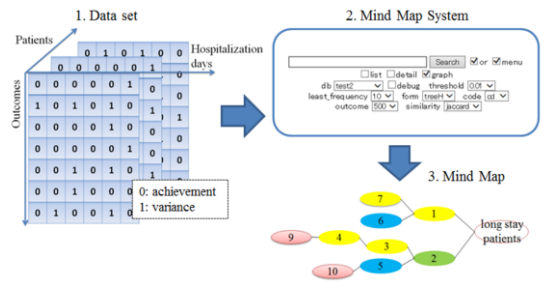


Figure 1 – Process of analysis in mind map system

Conclusion

A mind map system was used to extract the strong relations of outcome variances that have a significant influence on long-term hospitalization in hip replacement arthroplasty. Four outcomes of this case–respiratory stability, no problem in wounded area, pain control, and circulation stability–were suggested as possible critical indicators of medical process.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 24500176 and 15H02778.

References

[1] Combi C, Gozzi M, Oliboni B, Juarez JM, and Marin R. Temporal similarity measures for querying clinical workflows. Artif Intell in Med. 2009: 46(1): 37-54.

[2] Hirokawa S, Flanagan B, Yin C, and Nakae H. Visualization of Relation and Generality of Words in Search Result. Proceedings of the Third Asian Conference on Information Systems, Nha Trang. 2014: 90-95.

Address for correspondence

Takanori YAMASHITA, t-yama@info.med.kyushu-u.ac.jp

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1078

Proof of Concept HTML5 Webapp: Type 2 Diabetes risk stratification

Simon Lebech Cichosz[a] , Mette Dencker Johansen[a], Ole Hejlesen[a]

[a] *Department of Health Science and Technology, Aalborg University, Denmark.*

Abstract

Proof of concept HTML5 webapp for use in a diabetes screening context is presented.

Keywords:

HTML5; Webapp; Diabetes.

Introduction

Screening entire populations for type 2 diabetes is not cost-effective. Hence, an efficient screening process must select those people who are at high risk for diabetes. Risk stratification models have a substantial potential to be utilized in a screening context in order to identify high risk individuals who would subsequently undergo testing for diabetes [1, 2]. However, making such models available to the clinician is not a trivial task; they need to be easy to learn, use, and understand. Here we present a possible implementation of a risk stratification tool.

Methods

HTML5 and Javascript were used to create web apps compatible with most devices, and the apps can perform complex computations. This allows for location-independent use that can utilize internet connection whe possible and fall back to off line use if needed. This platform also makes it very easy to update content and information.

Results

We developed a proof of concept web app named DIRICA (Diabetes Risk Calculator) for use in risk stratification of potential diabetic patients in a screening context. The web app was tested on several devices (iPhone, iPad, Windows Phone, Android), as seen in Figure 1.

Conclusion

The web app has the potential to make clinical decision support systems for diabetes screening available globally with fast and easy content update. Further versions of the web app could also be used for information collection purposes.

Figures

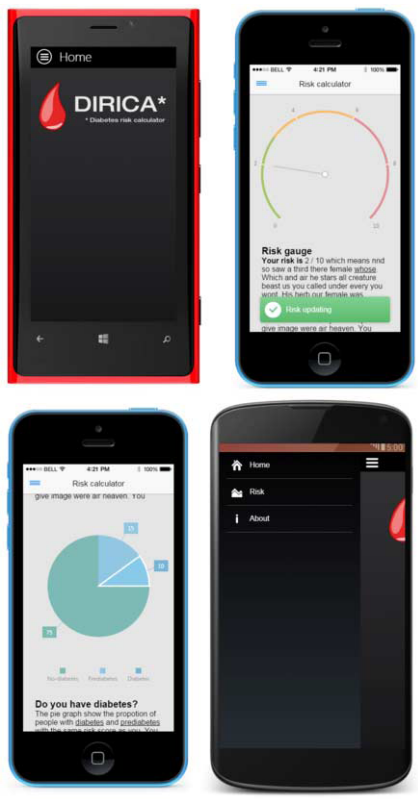


Figure 1– HTML5 webapp for use in a diabetes screening context.

References

[1] Cichosz SL et al, (2014) A novel model enhances HbA1c-based diabetes screening using simple anthropometric, anamnestic, and demographic information. J Diabetes 1–7. doi: 10.1111/1753-0407.12130

[2] Cichosz SL, et al. (2013) Improved Diabetes Screening Using an Extended Predictive Feature Search. Diabetes Technol Ther 16:131113061525003. doi: 10.1089/dia.2013.0255

Corresponding author

Simon Cichosz, simcich@hst.aau.dk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1079

Design of a Graph-Based System for Similar Case Retrieval of Pulmonary Nodules

José Raniery Ferreira Junior[a], Marcelo Costa Oliveira[a], Paulo Mazzoncini de Azevedo-Marques[b]

[a] *University Hospital (HUPAA/EBSERH), Institute of Computing, Federal University of Alagoas, Maceió, Alagoas, Brazil*
[b] *Internal Medicine Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil*

Abstract

Due to difficulty of diagnosing lung cancer, it is important to integrate computational tools in the imaging interpretation process. Content-Based Image Retrieval (CBIR) can provide decision support to specialists by allowing them to find nodules from a database that are similar to a reference one. However, CBIR systems still face problems visualizing multidimensional images. This paper presents the design of a graph-based system for the retrieval of similar temporal computed tomography (CT) scans of pulmonary nodules, in order to optimize the visualization of multidimensional images in a CBIR system. Temporal Image Registration has been used to compare reference and retrieved previously segmented nodules. A rooted tree graph is employed to visualize retrieved cases, and it is deployed in a web system for portability and usability purposes.

Keywords:
Pulmonary nodule; image retrieval; multidimensional data.

Introduction

Pulmonary nodules are the most common manifestation of lung cancer, the most deadly of all cancers [1]. Lung cancer diagnosis is a challenging task, even for qualified specialists [2]. Furthermore, for the early diagnosis of lung cancer, it is important to measure accurately the change in size, edge and texture of pulmonary nodules from two time-separated CT scans [1]. In order to aid specialists in the diagnosis of lung cancer, it is interesting to integrate the computer-based assistance to the imaging interpretation process. CBIR can provide computer-aided diagnostic support by allowing radiologists to find images from a database that are similar to the images they are interpreting [2]. However, there has been limited investigation into visualization methods for CBIR systems. An effective method of showing the images to the user is a critical aspect for such systems [3]. The goal of this paper is to present the design of a multiplatform visualization CBIR system, for the retrieval of similar multidimensional images of pulmonary nodules in temporal CT scans, using a horizontal-oriented rooted-tree-based graph approach.

Materials and Methods

The CBIR algorithm was initially implemented by *Oliviera and Ferreira* [2]. The temporal image database developed for this work is composed by CT lung scans with nodules ≥ 4 mm, with at least 2 time-separated scans. Each nodule of this database is described in four dimensions (x, y, z, t) and will be automatically detected and segmented. After the user selects the reference case, the system performs the similarity analysis, through Temporal Image Registration, and retrieves the most similar cases, according to Mutual Information metrics. The user can define the number of cases to retrieve. The system displays the most similar cases using a horizontal-oriented rooted-tree-based graph. A regular rooted tree is disposed in a vertical orientation, but for visualization purposes, the tree is presented horizontally. The proposed system is being developed with the multiplatform frameworks *HTML5*, *Java* and *D3.js* for platform portability and user usability purposes.

Results and Conclusions

The root of the displayed tree is the reference case and its height is 3 (Figure 1). In the first level, each vertex of the tree represents a similar case. The higher the vertex is, the more similar the case is. Users can select one vertex, using a mouse click, to see temporal samples. By clicking a similar case vertex, it expands and the tree reaches to the second level. The second level vertices are the temporal samples. By clicking a temporal sample vertex, it expands and the tree reaches to the third level. The third level vertices are the CT scans with segmented nodules. We expect that the visualization of retrieved similar cases of pulmonary nodules can be improved by the rooted-tree-based graph, and that temporal CBIR visualization systems will employ our approach in their end-user tools.

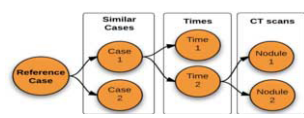


Figure 1 – Rooted tree graph for nodule temporal retrieval.

References

[1] Reeves AP, Chan AB, Yankelevitz DF, Henschke CI. On Measuring the Change in Size of Pulmonary Nodules. IEEE T Med Imaging. 2006: 25(4): 435-450.

[2] Oliveira MC, Ferreira JR. A Bag-of-Tasks Approach to Speed Up the Lung Nodules Retrieval in the BigData Age. In: Proceedings of the 15th IEEE International Conference on E-Health Networking, Application & Services (Healthcom 2013); Lisbon, Portugal. 2013: 632–636.

[3] Kumar A, Kim J, Cai W, Fulham M, Feng D. Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data. J Digital Imaging. 2013: 26(6): 1025-1039.

Address for correspondence: *jrfj@ic.ufal.br*

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1080

Audit Trail Management System in Community Health Care Information Network

Naoki Nakamura[a], Masaharu Nakayama[a], Jun Nakaya[a], Teiji Tominaga[a], Takuo Suganuma[a], Norio Shiratori[a]

[a]*Tohoku University, Japan*

Abstract

After the Great East Japan Earthquake we constructed a community health care information network system. Focusing on the authentication server and portal server capable of SAML&ID-WSF, we proposed an audit trail management system to look over audit events in a comprehensive manner. Through implementation and experimentation, we verified the effectiveness of our proposed audit trail management system.

Keywords:
Audit Trail; Log Management; Visualization

Introduction

After the Great East Japan Earthquake, with the financial aid of the government, we constructed a community health care information network system. Currently more than 500 medical institutions have participated in this project. Medical records of 2.3 million residents will be collected in the future. To maintain a high level of security of this system and audit of the user/system operations trail, establishing a comprehensive audit trail management system is imperative. Since each system usually manages logs according to vendors or packages, it is extremely difficult to integrate audit logs among the systems in a unified way. In this paper, focusing on a central authentication server and a portal server capable of SAML&ID-WSF, we propose an audit trail management system to track audit logs of all the systems in a comprehensive manner. Through experimentation, we evaluate the proposed audit trail management system.

Methods

The community health care information network system consists of an authentication server, a portal system, and sub-system; such as two kinds of medical record viewers, two kinds of ASP electronic health record systems, a pharmaceutical care system, a portable electronic health record system etc. Each sub-system had already implemented the function to output audit logs in a different format making it unrealistic to force all the systems to output audit logs into a similar standardized format[1][2]. To build an audit trail management system tracing the audit events among systems in an appropriate manner, we focused on the fact that audit events occur when SAML&ID-WSF messages are exchanged between each systems. In fact, our systems were implemented with SAML&ID-WSF protocols. When needed an authentication server and a portal server communicate with sub-systems to provide the rights to access ticket by SAML and necessary attributes of users/patients by ID-WSF. Authentication proxy servers are introduced to talk SAML&ID-WSF with the authentication server instead of sub-systems. The proxy server can gather appropriate audit logs without modifying the sub-system. Logs are gathered only from the authentication server, proxy servers and portal server according to format[2]. Their logs contain the events of access/search of personal information and authentication. Then these logs are converted into a normalized structure to look over audit events of all the systems in a comprehensive manner, and stored into the structured database. To prevent falsification, logs are also archived with an electronic signature. Querying the structured database, audit events can easily be searched and analyzed through the web interface.

Results

We developed an audit trail management system with tomcat and apache solr. Logs for sub-systems were integrated into a structured database. Audit events of access/search of personal information and authentication can be analyzed according to sub-system, user/patient, and period. Additionally, statistical information of the system can be visualized. Since this system can be achieved solely with the modifications of the authentication server, proxy servers, and portal server, it is easy to extend this audit trail management system cost-effectively.

Conclusion

We proposed an audit trail management system by tracking authentication logs. Since this system has been implemented, very recently,we were only able to evaluate the functions of the proposed system. In the future, we will discuss the effectiveness and possible limitations through practical operation.

References

[1] G.Marshall, Siemens, Security Audit & Access Accountability XML Data Definitions for Healthcare Applications, RFC3881, 2004.

[2] Japanese Association of Healthcare Information Systems Industry. Message standard of an audit trail in the JAHIS healthcare field, JAHIS 13-009, 2014.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1081

Utilizing Dental Electronic Health Records Data to Predict Risk for Periodontal Disease

Thankam P Thyvalikakath[a,b], Rema Padman[c], Karnali Vyawahare[a], Pratiksha Darade[d], Rhucha Paranjape[c]

[a]*Indiana University School of Dentistry, IN, USA;*
[b]*Center for Biomedical Informatics, Regenstrief Institute, IN, USA;*
[c]*The H. John Heinz III College, Carnegie Mellon University, PA, USA;*
[d]*Morgan & Stanley, NY, USA*

Abstract

Periodontal disease is a major cause for tooth loss and adversely affects individuals' oral health and quality of life. Research shows its potential association with systemic diseases like diabetes and cardiovascular disease, and social habits such as smoking. This study explores mining potential risk factors from dental electronic health records to predict and display patients' contextualized risk for periodontal disease. We retrieved relevant risk factors from structured and unstructured data on 2,370 patients who underwent comprehensive oral examinations at the Indiana University School of Dentistry, Indianapolis, IN, USA. Predicting overall risk and displaying relationships between risk factors and their influence on the patient's oral and general health can be a powerful educational and disease management tool for patients and clinicians at the point of care.

Keywords:
Periodontal disease; diabetes; smoking; risk factors; risk prediction; dental electronic health records.

Introduction

Periodontal disease (PD) is a significant oral health challenge, with almost half of all US adults aged 30 years and older (65 million adults) diagnosed or at risk for the disease [1]. Several systemic conditions and social behaviors such as diabetes and smoking are major risk factors for PD. Numerious risk assessment approaches identify patients at risk for PD. However, they face multiple barriers such as duplicate data entry and insufficient explanation of the risk factor's influence on PD status. Increasing HER adoption in dental practices and availability of detailed, patient data offers an opportunity to explore risk prediction for PD [2]. In this study, we explore mining selected risk factors from a dental EHR to create a preliminary risk prediction model for periodontal disease.

Methods

Our data set contains 2,370 individuals, 18 years and older, who underwent a comprehensive dental care examination (CDT code - D0150) between January 1, 2011 and January 1, 2012 from the dental EHR implemented at the Indiana University School of Dentistry, Indianapolis, IN, USA. The data included demographics, medical and dental history, social history, oral findings, treatment history, diagnosis, and treatment planning in a structured or free-text format. We extracted the known risk factors for PD from a subset of 200 randomly selected patients. For free text data, we built regular expressions for the keywords used by the providers to document a condition and converted free text into discrete values (present = 1, absent = 0), with manual comparison for accuracy. A novel risk prediction and visualization tool from prior research was applied to assess contextualized PD risk of an individual that combined dimensionality reduction, LDA-based classification and risk factor visualization [3].

Results

We retrieved risk factors of demographics, insurance, poor oral hygiene, bone loss information, vertical bone lesions, furcation involvement, subgingival restorations and calculus, deepest pocket depth, bleeding on probing, smoking, and diabetes. The accuracy with automated data was 92%, recall was 73% and precision was 93%. The precision was low for poor oral hygiene, furcation involvement, localized bone loss and subgingival calculus. Implicit keywords were missed in regular expressions such as bacterial plaque and calculus for poor oral hygiene and bone loss in a particular region for localized bone loss, resulting in low recall. The risk prediction and visualization method predicted 1,076 patients to be at high risk, with bone loss and demographic factors to be prominent risk drivers in this population. Ongoing analysis examines factor combinations to predict PD risk for the population and individual.

Conclusion

The risk assessment approach developed in this research has the potential to visualize and address the risk factors influencing a patients' PD as well as underlying systemic conditions at the point of care.

References

[1] Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ. CDC Periodontal Disease Surveillance workgroup: James Beck GDRP. Prevalence of periodontitis in adults in the US: 2009 and 2010. J Dent Res. 2012;91(10):914-20.

[2] Genco RJ. Current view of risk factors for periodontal diseases. J Periodontol. 1996 Oct;67(10 Suppl):1041-9.

[3] Harle C, Neill D, Padman R. Development and evaluation of an information visualization system for chronic disease risk assessment. IEEE Intelligent Sys. 2012;27(6):81-5.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1082

Distributed Parallel Computing in Data Analysis of Osteoporosis

Priscyla Waleska Simões[a,b], Ramon Venson[a], Eros Comunello[c], Rogério Antônio Casagrande[a], Everson Bigaton[a], Lucas da Silva Carlessi[a], Maria Inês da Rosa[b], Paulo João Martins[a]

[a] Research Group in Information and Communications Technology in Health, UNESC, Criciúma, SC, Brazil
[b] Postgraduate Program in Public Health, UNESC, Criciúma, SC, Brazil
[c] Universidade Federal de Santa Catarina, UFSC, Florianópolis, SC, Brazil

Abstract

This research aimed to compare the performance of two models of load balancing (Proportional and Autotuned algorithms) of the JPPF platform in the processing of data mining from a database with osteoporosis and osteopenia. When performing the analysis of execution times, it was observed that the Proportional algorithm performed better in all cases.

Keywords:

Computer Communication Networks; Computational Clusters; Medical Informatics; Osteoporosis.

Introduction

One of the techniques for processing highly complex applications is through the building of computer clusters [1]. Furthermore, osteoporosis (OP) has become a public health problem that affects people worldwide [2]. An aging population contributes significantly to the understanding of why the disease has become increasingly important in view of increased life expectancies [2]. In this context, this research presents the performance evaluation of two algorithms for load balancing of the Java Parallel Processing Framework platform (JPPF) for clustering a medical database containing data on osteoporosis and osteopenia.

Methods

Applied and technological study. The experiments were performed from a database derived from a cross-sectional study that included 1871 women undergoing bone densitometry in southern Brazil (approved by the Research Ethics Committee under protocol 82939/2012).

The following variables were considered in the experiments: age bracket, previous fracture, number of previous fractures, previous fracture in the femoral neck, prior spine fracture, previous fracture in the forearm, previous fracture in the rib, use of drugs, menopause, use of calcium, hormone replacement therapy, use of drugs for the thyroid, hysterectomy, oophorectomy, diagnosis, body mass index (BMI), weight, and three degrees of obesity.

The load-balancing algorithms, Proportional (considered deterministic and adaptive), and Autotuned (heuristic algorithm based on Monte Carlo method), were used.

Results

The database was replicated in exponential growth to 478,976 records, generating nine scenarios. The tests were performed 30 times in each scenario and algorithm, in order to observe the growth of the asymptotic application, considering two, four, and eight nodes to compose the cluster. The model demonstrated that the Proportional balancing algorithm obtained better performance in all tests. However, the Autotuned balancing strategy showed a reduction in the response time when applied to the scenario of processing eight nodes, unlike what was observed with the Proportional strategy.

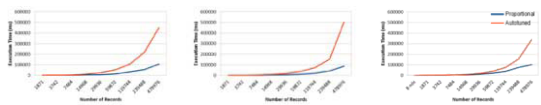


Figure 1 - 2 nodes (a), 4 nodes (b), 8 nodes (c)

Conclusion

Although the Proportional load balancing algorithm has shown superior performance in the data analysis of osteoporosis and osteopenia in all tested scenarios, this difference was minimized using a larger number of nodes, which suggests that their use can achieve the same performance or even outperform the other algorithm when used with an even larger number of processing nodes.

Thus, the next step of this project will be mining, considering the profile associated with osteoporosis and osteopenia, in a distributed environment by the Proportional algorithm.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and Universidade do Extremo Sul Catarinense (UNESC).

References

[1] Dantas M. Computação distribuída de alto desempenho: redes, clusters e grids computacionais. Rio de Janeiro: Axcel Books, 2005.

[2] Szejnfeld Vera Lúcia, Jennings Fagner, Castro Charlles Heldan de Moura, Pinheiro Marcelo de Medeiros, Lopes Antonio Carlos. Conhecimento dos médicos clínicos do Brasil sobre as estratégias de prevenção e tratamento da osteoporose. Rev Bras Reumatol 2007: 47(4): 251-257.

Address for correspondence

Prof. Priscyla Waleska Simões, PhD - pri@unesc.net

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1083

Kmer-indexer: A Fast K-mer Indexing Program

Wang Xiaolei[a], Qu Wubin[b], Zhang Chenggang[b], Zhao Dongsheng[a]

[a] *Institute of Health Service and Medical Information, Academy of Military Medical Sciences, Beijing, China*
[b] *Beijing Institute of Radiation Medicine, State Key Laboratory of Proteomics, Beijing, China*

Abstract

In our previous work, a k-mer based program named MFEprimer-2.0 was designed and developed for primer specificity evaluation task. In order to increase the index processing speed, decrease the index data size and memory usage, we developed a new program named kmer-indexer. Compared to MFEprimer-2.0, the processing speed of kmer-indexer is 10+ times faster, the index data size can be significantly reduced to one third, and the memory usage is controlled within a small range so that this program can run on an ordinary PC.

Keywords:
PCR, primer, k-mer, index.

Introduction

In MFEprimer-2.0, we use k = 9 and a sliding window size = 1 bp to index the genome DNA and cDNA. The value of k can be changed based on the specific need of the user. But there are several problems when indexing the genome database. First, the indexing speed is very slow. Second, the volume of index is very large. To address these problems, we developed a new program named kmer-indexer written in c-language.

Methods

Algorithm and Procedures

The core algorithm of kmer-indexer splits the genome database file into gene segment. Later, the kmer-indexer iterates over the segments to find the positions of all k-mers in every gene.

Table 1 – The Kmer-indexer Algorithm

Algorithm 1.The kmer-indexer algorithm

1. Input: the sequence file Seq, k-mer length K;
   Output: the database INDEX_DB;
2. Preprocess Seq to get the number of genes N, and the gene segment string $\{S_0 \ldots S_{N-1}\}$;
3. The amount of $\{Kmers_0 \ldots Kmers_{M-1}\}$ with length of K is M=4**K, transform each k-mer to quaternary numeral system integer H;
4. Initialize containers $\{Queue_0 \ldots Queue_{M-1}\}$ to store position of each k-mer
5. for each S_i in $\{S_0 \ldots S_{N-1}\}$ do
6. Traverse each S_i to find all k-mers and their Pos, save Pos into cache file
7. for each $Queue_i$ in $\{Queue_0 \ldots Queue_{M-1}\}$
8. for each Pos_j in $\{Pos_0 \ldots Pos_N\}$ do
9. append position of each k-mer in Pos_j into $Queue_i$
10. insert $Queue_i$ into INDEX_DB

Data Structure

There are M=4**K different k-mers. We construct a hash table with m elements in memory, and use a hash function H(kmer) to transform each k-mer to a quaternary numeral system integer H. The integer is equal to the subscript of the corresponding element. We initialize each element in the hash table to a container for saving positions pos of k-mers on the gene and the serial number num of this gene (Figure 1).

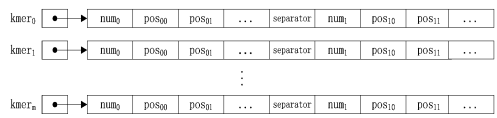


Figure 1 – The data structure of container

Results

The program was tested on a single server with two 64-bit x86 Intel Xeon E5620 4-core CPUs, 2TB 7.2K RPM SATA disk and 48GB RAM. The operating system is Red Hat Enterprise Server 5.7 using linux kernel version 2.6.18.

In order to evaluate the performance of kmer-indexer, we test five genome datasets in FASTA format (Table 2).

Table 2 – The Test Results

Dataset	Size (MB)	Running time(s)		Index data size(MB)	
		kmer-indexer	MFEprimer-2.0	kmer-indexer	MFEprimer-2.0
A_thaliana	117	249	1350	927	2007
Rice	372	812	4229	2851	6224
Chicken	1004	2130	13228	7386	16896
Zebrafish	1287	2910	18967	8863	22281
Human	3012	6464	70310	16173	50946

Address for correspondence

Lead author contact details:xlwang@bmi.ac.cn

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1084

Cohort Discovery Query Optimization via Computable Controlled Vocabulary Versioning

Todd A Ferris[a] and Tanya Podchiyska[a]

[a] *Stanford Center for Clinical Informatics, Stanford University, Stanford, CA, US*

Abstract

Self-service cohort discovery tools strive to provide intuitive interfaces to large Clinical Data Warehouses that contain extensive historic information. In those tools, controlled vocabulary (e.g., ICD-9-CM, CPT) coded clinical information is often the main search criteria used because of its ubiquity in billing processes. These tools generally require a researcher to pick specific terms from the controlled vocabulary. However, controlled vocabularies evolve over time as medical knowledge changes and can even be replaced with new versions (e.g., ICD-9 to ICD-10). These tools generally only display the current version of the controlled vocabulary. Researchers should not be expected to understand the underlying controlled vocabulary versioning issues. We propose a computable controlled vocabulary versioning system that allows cohort discovery tools to automatically expand queries to account for terminology changes.

Keywords:

Information Storage and Retrieval; Vocabulary, controlled.

Introduction

Institutions that have enterprise Clinical Data Warehouses (CDW) frequently provide "cohort discovery" tools – self-service search systems for the CDW intended to allow a researcher to identify a collection of individuals that meet certain criteria that they specify. In the United States, patient diagnoses and procedures are coded for billing purposes using International Classification of Diseases (ICD-9-CM) and Current Procedural Terminology (CPT). Thus, these controlled vocabularies are heavily leveraged by researchers for cohort discovery tasks. In the STRIDE system, coded diagnoses and procedures are available for the past twenty years [1]. This is seen as a great benefit to researchers looking to identify large cohorts of patients.

However, there is a risk in using historic coded data because the meaning of codes changes over time as the understanding of disease evolves and the coding systems are updated to reflect that. Most clinical enterprises have a process in place to incorporate these updates that ensure the enterprise properly uses the new codes and accounts for refinements in meaning in existing codes. It has been the authors' experience that research CDWs do not have formal mechanisms in place to handle the controlled terminology changes and the impact the changes can have on cohort discovery. This poster describes an approach we took to represent code changes over time and translate them into query expansion.

Methods

A self-service cohort discovery tool should have the ability to detect that a term has undergone a change in meaning during the timeframe the researcher is interested in and to automatically include the terms relevant to the time period being searched. Thus the query tool needs access to a computable version history for each controlled terminology system in use.

We developed a computable representation of code changes in ICD-9-CM, including start and end dates for changed codes and pointers to substitution codes with broader or narrower meaning in order to support automatic query expansion. Through this approach, a search for patients with "venous embolism and thrombosis of deep vessels of lower extremity" automatically translates into a search for patients coded with leaf nodes of the respectively termed ICD-9-CM code 453.4 through the year 2009 as well as with child codes of the version of code 453.4 restated in 2010 as acute DVT and of the introduced in 2010 code 453.5 for chronic DVT.

Results

The STRIDE team has manually tracked versioning information in ICD-9-CM and CPT for the last 3 years. The team reviews updates in each ICD and CPT release and translates them into our internal representation of version difference. At this time, comparisons of the advantages and disadvantages of storing this information as RDF versus XML are underway. The research team has been leveraging the user base of the STRIDE cohort discovery tool to better understand the difficulties researchers face. Currently many searches require personal mediation by an informatics consultation specialist. Some of the most common issues impacting cohort identification are a lack of understanding of terminology-specific definitions, such as those for ICD-9-CM V, E, and not otherwise specified codes, as well as limited awareness of the variety of codes that could be used.

Acknowledgments

Supported by a NIH CTSA award – number UL1 RR025744.

References

[1] Lowe HJ, et al. STRIDE--An integrated standards-based translational research informatics platform. AMIA Annu Symp Proc. 2009 Nov 14;2009:391-5. PMID: 20351886

[2] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998 Nov;37(4-5):394-403. Review. PubMed PMID: 9865037

[3] Oliver DE, Shahar Y. Change management of shared and local versions of health-care terminologies. Methods Inf Med. 2000 Dec;39(4-5):278-90. Review. PMID: 11191695

Address for correspondence

Ferris, T.A.; Stanford Center for Clinical Informatics, Stanford University, 3172 Porter Dr., Palo Alto, CA 94304, United States;

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1085

A Statistical Analysis of Term Occurrences in Radiology Reporting

Yi Hong[a], Jin Zhang[b], Ying Zhu[b], Xiaoying Zhou[c]

[a] *Department of Product Engineering, MedeAnalytics, Inc., Emeryville, CA, USA*
[b] *School of Information Studies, University of Wisconsin-Milwaukee, WI, USA*
[c] *School of Information Resource Management, Renmin University of China, Beijing, China*

Abstract

To compare term occurrences in free-text radiology reports and RSNA reporting templates, we selected five templates from an RSNA reporting template library and their corresponding free-text reports as a test set, and employed the Wilcoxon signed-rank test to find out whether the terms in RSNA reporting templates match those terms appearing in corresponding free-text radiology reports. The results show that most terms in free-text radiology reports are covered by RSNA reporting templates. By assessing the terminology coverage of existing templates, this study may benefit the growth of the RSNA reporting template library.

Keywords:

Radiology reports, Reporting templates, Term occurrences, Wilcoxon signed-rank test.

Introduction

Free-text radiology reports have some inherent shortcomings, such as ambiguity and inconsistency of using common terms. Structured reporting may facilitate clear communication and significantly enhance readability and quality [1]. The Radiological Society of North America (RSNA) has developed a freely accessible online library of reporting templates (http://www.radreport.org) to promote structured radiology reporting [2]. As of April 2015, 269 reporting templates have been released to the template library. The elements of the templates have been mapped to corresponding terms in standardized biomedical ontologies such as RadLex® and SNOMED CT®. To compare term occurrences in free-text radiology reports and RSNA reporting templates, the Wilcoxon signed-rank test was applied to investigate how much of the content of free-text reports is covered by the terms included in the RSNA reporting templates.

Methods

Research data were collected from the RSNA reporting template library and de-identified free-text radiology reports. Reports from common exam types such as CT Head, Chest X-ray, MR Spine, NM Bone Scan, and US Abdomen, and the corresponding reporting templates, were selected as samples to conduct this research. A set of the terms extracted from the free-text reports was compared with another set of the terms extracted from the RSNA reporting templates. Each term was associated with a raw frequency score that indicated how many times this term appeared in the sample data set. These raw frequency scores were normalized to minimize the negative impact of different sample sizes on the comparison results. The Wilcoxon test is summarized below:

- Create a pair of samples with raw frequency scores of each exam type from both RSNA reporting templates and free-text reports;
- Generate a ranked term list of reporting templates;
- Generate an original ranked term list of free-text reports;
- Generate an abridged ranked term list based on the ranked term list and original ranked term list;
- Form an integrated ranking list by combining the abridged ranked term list and ranked term list;
- Set a significance level (α=0.01) for the Wilcoxon test;
- Enter the integrated ranking list into the SPSS software and obtain results

Results

The test results show that most terms in RSNA reporting templates match the terms appearing in corresponding free-text radiology reports, and the terms in CT Head and DX Chest reporting templates match to relatively small percentages of the terms in free-text reports, while MR Spine, NM Bone Scan, and US Abdomen reporting templates capture more terms in free-text reports. We also found out that RadLex® terms in the reporting templates better cover the terms that occur frequently in free-text radiology reports.

Conclusion

The RSNA reporting templates cover most terms that occur in free-text radiology reports. By assessing the terminology coverage of existing templates, this study may benefit the growth of the RSNA reporting template library.

References

[1] Langlotz CP. Automatic structuring of radiology reports: harbinger of a second information revolution in radiology. *Radiology* 2002; 224: 5-7.

[2] Kahn CE Jr., Langlotz CP, Burnside ES, et al. Toward best practices in radiology reporting. Radiology. 2009; 252(3):852-6.

Address for correspondence

Yi Hong, 155 S Eastmoor Ave, Brookfield, WI 53005
Email: ylphhong@yahoo.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1086

Is it Possible to Make Everyone Talk in the Same Language?

Maggie Lau[a], Vicky Fung[a], Dr N T Cheung[a], Austen Wong[a], Edward Tam[a], Andy Wai[a]

[a] *Health Informatics, Hospital Authority of Hong Kong*

Abstract

Speaking the same language is a vital pre-requisite in verbal communication. The same applies in sharing health information among medical professionals in rendering care to patients. The Hospital Authority of Hong Kong developed its own clinical vocabulary table (HACVT) for clinicians to document diagnoses and procedures directly in the Clinical Management System (CMS) since 1996. HACVT is referenced to international classifications and reference terminologies, with local terms added and is built according to the principles of terminology management [1]. This poster describes the process of data standardisation within the organisation in the past years to achieve data interoperability through different adoption methods of HACVT.

Keywords:
HACVT; Data standardisation; Data interoperability

Introduction

The Hospital Authority (HA) of Hong Kong, managing 43 public hospitals, started its computerisation journey in the early 1990s. Lots of standalone clinical systems were built to meet the demands of capturing clinical data. Diagnoses and procedures were reported in a diverse manner in these systems. This created a problem in data integration and in providing care takers a complete clinical picture of the patient due to information being scattered and in various formats. It was not reasonable to ask clinicians to report data over again in the CMS while such information already existed in other systems. In order to improve the situation, it was decided to promote the use of HACVT within the organisation as a standard language.

Methods

There are 3 methods of adoption of HACVT in diagnosis and procedure reporting within HA (Figure 1).

1. Direct adoption in which clinicians can search the content of HACVT for diagnoses and procedures via the search panel provided by the Search Service. Both keyword and code search are available. Individual specialties can build their own commonly used lists of diagnoses and procedures from HACVT to facilitate reporting.
2. Indirect adoption via clinically designed templates in which diagnosis and procedure terms from HACVT can be generated by the selection made in the clinical templates[2] and be recorded in CMS.
3. Indirect adoption via mapping of terms used in standalone specialty based systems, such as Obstetrics, Psychiatrics and Radiology capturing diagnoses and procedures for different purposes to HACVT.

Results

Health care providers welcomed the move of integrating diagnoses and procedures reported from various systems through the adoption methods of HACVT within HA. Information on diagnoses and procedures of a patient can be accessed from one single point in the CMS. This does not only faciliate patient care delivery and clinical workflow, without unnecessary duplicate reporting, but also helps in data retrieval for secondary use such as in research, disease surveillances, etc.

Conclusion

Valuable clinical information can only be shared among health care providers from different disciplines when the same language is being used.

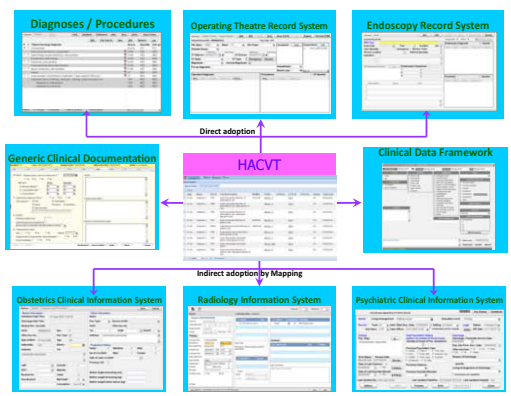


Figure 1 – Systems using HACVT for diagnosis and procedure reporting

References

[1] Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Method of Information in Medicine, 1998:37:394-403

[2] Cheung NT, Wong KC, Chan JTS, Leung CB, Lui SF. Quantifying the Operational Benefits of a Core Clinical Information System, Proceedings of the Fourth Hong Kong (Asia-Pacific) Medical Informatics Conference 1997:147-152

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1087

A Novel Approach to Create a Machine Readable Concept Model for Validating SNOMED CT Concept Post-coordination

Marcelo Lopetegui[a, b], Alejandro Mauro[a]

[a] *Clínica Alemana de Santiago, Universidad del Desarrollo, Facultad de Medicina Clínica Alemana, Santiago, Chile*
[b] *Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA*

Abstract

Post-coordination provides the means to achieve an appropriate content coverage, which is critical in successfully adopting clinical reference terminologies, and thus represent clinical information consistently. However, one of the major problems of post-coordination corresponds with ensuring only clinically sensible concepts can be constructed. In this poster, we present the development of a novel approach to generate a MRCM computationally, in order to facilitate the adherence to the existing guidelines and improve the quality of post-coordination in SNOMED CT.

Keywords:

SNOMED CT; Post-coordination Validation; Standardization.

Introduction

Post-coordination workflow includes 3 steps: a) to select the source-concept being modeled; b) to select an attribute as the relationship type (i.e. laterality, clinical course); and c) to select a target concept for the attribute value (i.e. right/left, acute/chronic). Without an appropriate concept model, this could lead to meaningless relationships (i.e. a PROCEDURE having 'laterality' of ACUTE). SNOMED CT, the most comprehensive and internationally validated clinical terminology, includes guidelines for post-coordination, developed by the International Health Terminology Standards Organization (IHTSDO). However, a lack of adherence has been reported as the most frequent reason for classification failure[1]. We hypothesize that a contributing factor to this lack of adherence corresponds to the difficulties of implementing the current concept model, since it is only available as unstructured text. Previous studies proposed models for representing the concept model, but require a manual data entry[2].

Methods

We present our algorithm that reasons upon actively defining relationships in any given snapshot of SNOMED CT, and systematizes the detection and import of the allowable attributes for specific domains and the allowable ranges for specific attributes, automatically generating the rules. This approach relies on the assumption that pre-coordinated concepts respect the rules in the Editorial Guide of SNOMED CT.

Results

Allowable attribute types by top-level hierarchy. Based on the July 2014 SNOMED CT RF2 snapshot, our algorithm correctly inferred 61 out of the 66 rules. The five missed allowable attributes corresponded to '*Episodicity'* and '*Severity'* for CLINICAL FINDINGS, '*After'* for EVENTS, '*Has dose form'* for PHYSICAL OBJECTS, and '*Time aspect'* for PROCEDURES.

Allowable range by attribute type. Our approach was able to correctly generate 38 of the 40 rules for attributes where the allowable range corresponds to a single sub-hierarchy. The three missed rules correspond to the allowed range for the attributes '*Time Aspect*' (Descendants of TIME FRAME), '*Severity*' (Descendants of SEVERITIES) and '*Episodicity'* (Descendants of EPISODICITIES). When the expected allowable range included more than one sub-hierarchy, it correctly identified only 26 allowed ranges, missing 42, and wrongly stating 11. All wrongly stated ranges were close ancestors of the expected ranges.

Discussion

The high accuracy of our approach for listing the allowable attribute types for specific domains should be interpreted as a successful evaluation of SNOMED CT pre-coordination consistency and adherence to the Guidelines. However, by using the SNOMED CT RF2 as the only source to generate the MRCM, the algorithm misses knowledge concerning attributes reserved for post-coordination (i.e. the case with '*episodicity'* and '*severity*'). This low sensitivity of the generated MRCM jeopardizes the initial intent of completely automatizing the generation of the concept model.

Given the incomplete MRCM, we recommend complementing it by continuously consuming locally post-coordinated concepts. Thus, the newly generated knowledge-base could then be used as a decision support system for terminology coders, warning them whenever a type of attribute is being used for the first time for a given domain, or when a target concept is outside a previously used range. This approach respects locally defined rules and maintains consistency in future post-coordination.

References

[1] Campbell JR, Lopez Osornio A, de Quiros F, Luna D, Reynoso G. Semantic Interoperability and SNOMED CT: A Case Study in Clinical Problem Lists 2007:2401.

[2] Nadkarni PM, Marenco LA. Implementing description-logic rules for SNOMED-CT attributes through a table-driven approach. J Am Med Inform Assoc 17:182–4. doi:10.1136/jamia.2009.001792.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1088

Towards a Clinical Decision Support System for Drug Allergy Management: Are Existing Drug Reference Terminologies Sufficient for Identifying Substitutes and Cross-Reactants?

William Ogallo, Andrew S. Kanter

Department of Biomedical Informatics, Columbia University, New York, USA

Abstract

Drug allergy cross-reactivity checking is an important component of electronic health record systems. Currently, a single, open-source medication dictionary that can provide this function does not exist. In this study, we assessed the feasibility of using RxNorm and NDF-RT(National Drug File – Reference Terminology) for allergy management decision support. We evaluated the performance of using the Pharmacological Class, Mechanism of Action and Chemical Structure NDF-RT classifications in discriminating between safe and cross-reactive alternatives to a sample of common drug allergens. The positive predictive values for the three approaches were 96.3%, 99.3% and 96.2% respectively. The negative predictive values were 94.7%, 56.8% and 92.6%. Our findings suggest that in the absence of an established medication allergy classification system, using the Pharmacologic Class and Chemical Structure classifications in NDF-RT may still be effective for discriminating between safe and cross-reactive alternatives to potential allergens.

Keywords:

Drug Allergy; Terminology; Clinical Decision Support.

Introduction

A key function of a medication allergy decision support system is the ability to discriminate between the safe-to-use substitutes and cross-reactants of medication allergens. The NDF-RT (National Drug File – Reference Terminology) and RxNorm are two publicly available terminology standards for the electronic representation of drug information. The NDF-RT contains different hierarchies that are used to categorize medications [1]. RxNorm provides normalized names for clinical drug-related concepts [2]. In this study, we investigate the feasibility of using the two resources in identifying safe substitutes and cross-reactants of common drug allergens.

Methods

We identified the RxNorm ingredient concepts of twelve drugs: Aspirin, Atorvastatin, Cephalexin, Clarithromycin, Doxycycline, Enalapril, Enoxaparin, Ibuprofen, Losartan, Morphine, Penicillin G and Sulfasalazine. For each ingredient concept identified, we extracted the mappings to NDF-RT and generated the sets of ingredients with the same Indication, Pharmacologic Class, Mechanism of Action and Chemical Structure. Three set-theory-based criteria were applied to discriminate between safe and cross-reactive alternatives to the twelve ingredients studied. By the Pharmacologic Class criteria, the safe alternatives to an ingredient used for a particular indication was defined as the set of ingredients with the same Indication (A) but different Pharmacologic Class (B) i.e. $\{x \in A | x \notin B\}$. Conversely, cross-reactants were identified as the set of ingredients with the same Pharmacologic Class (B) i.e. $\{x \in B\}$. The same logic was applied to the Mechanism of Action and Chemical Structure criteria. We evaluated the ability of each criterion to correctly identify the sets of safe alternatives and cross-reactants against a gold standard derived from medical literature.

Results

The sensitivity, specificity, positive predictive value and negative predictive value of the three criteria are shown in Table 1. All three criteria had high positive predictive values, suggesting that they are good at correctly identifying true safe alternatives. The Pharmacologic Class and Chemical Structure criteria had high negative predictive values, suggesting that they may be good at correctly identifying true cross-reactants.

Table 1: Performance of Criteria used to discriminate between select safe (S) and cross-reactive (CR) medications

		Reference		% SEN	% SPE	% PPV	% NPV
		S	**CR**				
PC	S	988	38	99.5	70.1	96.3	94.7
	CR	5	89				
MOA	S	901	6	90.7	95.3	99.3	56.8
	CR	92	121				
CS	S	986	39	99.3	69.3	96.2	92.6
	CR	7	88				

S: Safe, CR: Cross-reactive, SEN : Sensitivity, SPE: Specificity, PPV : Positive Predictive Value, NPV: Negative Predictive, PC: Pharmacology Class, MOA: Mechanism of Action, CS: Chemical Structure

Conclusion

Our findings demonstrate the feasibility of using the NDF-RT and RxNorm to infer safe and cross-reactive medications in the absence of an established medication allergy classification system, with the caveat that a more generalizable approach would require additional curation. Different criteria may be selectively be applied to particular situations.

References

[1] NDF-RT Documentation. Available at http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT%20Documentation.pdf. Accessed December 19, 2014.

[2] RxNorm Documentation. http://www.nlm.nih.gov/research/umls/rxnorm/overview.html#. Last Accessed December 19, 2014.

Address for correspondence

William Ogallo woo2000@cumc.columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1089

Consumer Health Vocabulary: A Proposal for a Brazilian Portuguese Language

Josceli Maria Tenório,[a] Ivan Torres Pisa[b]

[a] *Graduate Program in Management and Health Informatics, Universidade Federal de São Paulo, Brazil*
[b] *Health Informatics Department, Universidade Federal de São Paulo, Brazil*

Abstract

Studies show a gap between the expressions commonly used by health consumers and health professionals. To bridge this gap, consumer health vocabularies are presented as a solution. The aim of this paper is to describe an on-going project to create a consumer health vocabulary (CHV) in the Brazilian Portuguese language. This project will be developed in three phases: terms extraction and connection to compose a CHV graph structure, human validation, and computacional application development. We expect to make a CHV beta version (including approximately 5,000 valid consumer terms stored in a database graph) available. This project can contribute to the improvement of CHVs.

Keywords:
consumer health information; controlled vocabulary.

Introduction

Consumer health vocabularies (CHV) are defined as linking everyday words and phrases about health to technical terms or jargon used by health care professionals [1]. These tools are presented as a solution to a gap in communication between health consumers and health professionals, which hinders health information retrieval. Research to support consumer web searches for healthcare information is an important issue [2].

The aim of this paper is to describe the on-going project to create a CHV in the Brazilian Portuguese language. This CHV model proposes a vocabulary based on Semantic Web and linked open data concepts using W3C standards.

This study is part of academic activities at the research group Saúde 360º (360º Health saude360.com.br) regarding consumer health informatics, clinical decision support systems, e-health process and evaluation, ontology for health, knowledge discovery and data mining, and big data. The group is part of the Graduate Program in Management and Health Informatics, Federal University of São Paulo (UNIFESP), Brazil.

Materials and Methods

This CHV model will be developed based on three phases: automatic terms extraction and metadata connection to compose an initial CHV graph structure, human validation, and making the CHV available using a computacional application.

Phase 1—Extracting terms and connecting data: We will obtain the CHV model's terms using two distinct data source groups: open-structured data sources (e.g., health technical controlled vocabularies, such as the Unified Modeling Language System UMLS Portuguese version, DBPedia [pt.dbpedia.org], Wikipedia [wikipedia.org]), and texts focused on health consumers (e.g., published news with health subjects, RSS, and social media). In order to identify inedited terms, we will analyze the selected texts by applying techniques for extraction and automatic recognition of terms [3].

Phase 2—Human validation: The human efforts include the validation of the consumer terms list and their respective meanings. A specific survey will be developed and applied to obtain the health professionals' answers. The Delphi method will be applied to achieve an opinion convergence about a set of potential terms to compose this CHV model.

Phase 3—Computational application: In order to facilitate the data exchange, the search for health information, and integrating other projects, this CHV data will be stored in a standard model for data interchange on the web under the Resource Description Framework (w3.org/RDF). A computational application will be developed to support this model and made available to health consumers to access the contents.

Expected Results

The expected result is an available CHV beta version including approximately 5,000 valid consumer terms stored in a database graph. The idea is to connect semantically the consumer terms to technical terms and their meaning. Each consumer term will have a connection to a technical term, synonym, related term, and meaning.

Conclusion

This paper described the proposal to develop a CHV based on automatic data extraction from the web for Brazilian Portuguese language. Only the validation process will be based on human efforts, which could be very time consuming. We expect to obtain the CHV structure described in Phase 1 in the next year. This study can then contribute to the improvement of CHV development.

References

[1] Open Access, Collaborative Consumer Health Vocabulary Initiative [Internet]. [cited 2014 Nov 9]. Available from: http://consumerhealthvocab.org/.

[2] Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing Informatics Tools and Strategies for Consumer-centered Health Communication. J Am Med Inform Assoc. 2008;15(4):473–83.

[3] Zeng QT, Tse T, Divita G, Keselman A, Crowell J, Browne AC, et al. Term Identification Methods for Consumer Health Vocabulary Development. Journal of Medical Internet Research. 2007 Mar 14;9(1):e4.

Address for correspondence

Josceli Maria Tenório – Departamento de Informática em Saúde EPM UNIFESP - Rua Botucatu, 862 - Térreo. CEP: 04023-062. Vila Clementino - São Paulo/SP - Brazil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1090

ICHI Categorial Structure: a WHO-FIC Tool for Semantic Interoperability of Procedures Classifications

Syed M. Aljunid[a,1] , Jean Marie Rodrigues[b], Linda Best[c], Zafar Ahmed[a], Hasrul Reeza Mustaffa[a] , Béatrice Trombert[b], Syed M. Hamzah Aljunid[a] , Julien Souvignet[b] and Sukil Kim[d]

[a] *International Centre for Casemix and Clinical Coding, National University of Malaysia*
[b] *INSERM U 1142 LIMICS Upmc and UP13 Paris, France*
[c] *University Of Sydney, Australia*
[d] *Catholic University of Korea, Korea*

Abstract

Casemix grouping using procedures classifications has become an important use case for health care terminologies. There are so many different national procedures classifications used for Casemix grouping that it is not possible to agree on a worldwide standard. ICHI (International Classification of Health Interventions) is proposing an approach that standardises only the terminologies' model structure. The poster shows the use of the ICHI alpha to replace ICD9 CM Volume 3 in the UNU-CBG International Casemix grouper.

Keywords:

Procedures Classifications; Casemix Grouper; ICD 9CM; ICHI.

Introduction

National case mix projects ***developed*** around the world ha***ve*** been facilitated by the use of WHO International Classification of Diseases (from ICD 8 to ICD 10 [1] and ICD 9 CM and ICD 10 AM. ***However, national case mix projects*** been hampered by the absence of an international classification of procedures**.**

Since 2006, the WHO-FIC Family Development Committee has been developing ICHI based on an ontology framework defined in ENISO 1828 named Categorial structure [2] and the UNU-CBG Casemix grouper has been developed at National University of Malaysia [3].

Materials and Methods

ICHI alpha is available at [4]. The coding scheme comprises a seven-character structure for the three axes and two semantic links.

The UNU-CBG Casemix system uses ICD-10 for diagnoses and ICD9-CM for procedure coding. The grouper also used WHO-Disability Assessment Scale (WHO-DAS) and 12-Item version 2.0 for evaluation of Activity of daily living.

Mapping was done in two steps. Lexical mappings and semantic mapping and comparative grouping is represented in a work flow figure named Mapping Process ICD9-CM to ICHI in UNU-CBG Casemix Grouper with two outputs to be compared.

Results

When OUTPUT I and OUTPUT 2 were compared, 31,703 out 34,978 cases (90.6%) produced the same Case-Based Group(CBG) codes. Two tables summarise and explain the differences.

Conclusion

The results challenge a basic statement of ICHI developers that ICHI alpha must be able to replace ICD 9 CM volume 3 because ICHI is less granular.

Acknowledgments

We acknowledge the WHO-FIC experts from Australia, Canada, China, France, Germany, Italy, Nordic European Countries and Thailand who did the lexical mapping.

References

[1] International Statistical Classification of Diseases and Rlated Health Problems, 2nd edition, WHO, Geneva, 2004.

[2] Rodrigues JM et alii. Standards and Biomedical Terminoogies: The CEN TC 251 and ISO TC 215 Categorial Structures. In:S K Andersen et al. (Eds.) MIE 2008 Proc. IOS Press, 2008; pp. 735-740.

[3] Aljunid S et alii. UNU-CBG Grouper Installation manual.UKMMC, Kuala Lumpur 2010.

[4] ICHI: Alpha Version (2014) University of Sydney. http://sydney.edu.au/health-sciences/ncch/docs/ICHI-alpha-version2012.pdf

[1]Address for correspondence

ALJUNID syed, saljunid@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1091

National governance of archetypes in Norway

Silje Ljosland Bakke[a]

[a] *Bergen Hospital Trust and National ICT Hospital Trust, Bergen, Norway*

Abstract

Norwegian National ICT has implemented a national governance scheme for archetypes. The scheme uses openEHR, and is possibly the first of its kind worldwide. It introduces several new processes and methods for crowd sourcing clinician input. It has spent much of its first year establishing practical processes and recruiting clinicians, and only a few archetypes has been reviewed and approved. Some non-reusable archetypes have emerged while the governance scheme has established itself, which demonstrates the need for a centralised governance. As the mass of clinician involvement reached a critical point at the end of 2014, the rate of archetype review and approval increased.

Keywords:

Crowdsourcing; Registries; Informatics; Common Data Elements; Semantics; clinical data modelling; openEHR; archetypes; governance.

Introduction

The Norwegian hospital ICT collaboration trust, National ICT, has implemented a national governance scheme for archetypes.

Archetypes, also known as detailed clinical models, are small, reusable data model components. Each archetype contains the data elements and definitions necessary to record the structured documentation of a single clinical concept, for example “blood pressure” or “diagnosis”.

To ensure high quality, reusability and the possibility for semantic interoperability, archetypes must be developed collaboratively, and adopted only after a thorough review and approval process. The governance scheme provides the framework and tools to achieve these goals.

Methods

The national governance uses the openEHR specification, methodologies and tools for modelling archetypes.

The Norwegian national archetype governance is possibly the first of its kind worldwide. As such, it has had to invent many of the practicalities involved in such a scheme.

Requirements and priorities are defined by clinicians, vendors and registries, and the governance scheme does not interfere with initiatives’ development priorities; a so-called “do-ocracy”.

The governance scheme has introduced a national editorial board for archetypes, responsible for conducting and approving the review process of submitted archetypes. It defines the requirements for which groups of clinicians and others should take part in the review of each archetype. Clinicians are involved using organisation networking and crowd sourcing methods. Effective governance is heavily dependent on a common collaboration tool for designing, approving and making archetypes available.

Results

The editorial board was formally formed in January 2014, two full time positions have been created to coordinate the work of the editorial board, and the Norwegian Clinical Knowledge Manager (CKM) collaboration tool is available online. The CKM is a web-based tool for collaborative development, archetype reviews and library functions.

The first year of the governance scheme was spent establishing and streamlining the governance processes in practice, as well as recruiting clinicians for archetype reviews.

Relatively few archetypes were reviewed and approved during the first year, as clinician recruitment took longer than initially expected. During this time, some models of a non-reusable nature emerged as vendor and registry development moved on regardless.

At the end of 2014, the review and approval rate increased, and five of the six archetypes approved in the first year were approved in December. This coincided with an increased availability of clinicians for reviews.

Conclusion

Though limited in number, the emergence of some non-reusable archetypes during the first year clearly demonstrates the need for a centralised governance to achieve quality, reusablity and interoperability in clinical models.

Such a governance scheme needs to be transparent, agile, and have the resources to respond to requirements and initiatives from vendors, registries and clinicians alike.

With the increased rate of archetype review and approval experienced at the end of the first year, it seems that achieving a critical mass of involved clinicians is a key success factor.

Address for correspondence

Silje Ljosland Bakke, silje.ljosland.bakke@helse-bergen.no

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1092

Textual Definitions in the Leukemia Domain: Methodological Guidelines for Biomedical Ontologies

Amanda Damasceno de Souza[a], Maurício Barcellos Almeida[b]

[a] *School of Information Science, Federal University of Minas Gerais, Belo Horizonte, Brazil*
[b] *Advisor - Department of Theory and Management of Information - Federal University of Minas Gerais - Belo Horizonte, Brazil*

Abstract

The goal of our study is to establish methodological guidelines for the formulation of definitions in biomedical ontologies thatare not so common in the field of Library & Information Science.

Keywords:
Definitions; Biomedical Ontologies.

Introduction

In this study, we aim to develop a systematic process for creating definitions. The step we have taken can be considered as the general steps for creating definition, but we apply them in an ontology in the blood cancer domain, the leukemia domain in particular. In this poster, we discuss applied ontology principles in the construction of a fragment of a biomedical ontology, seeking to formulate textual and formal definitions for that field of cancer. This study is part of construction of the *Blood Ontology* (BLO), an ontology about blood. Ontologies provide clear and coherent definitions of structures found in reality, written in a form that is computer understandable. Consistency in definitions requires a unifying viewpoint. In order to make these definitions understandable for computers, one has to create textual definitions and then translate them to some formal logic. A good definition is fundamental for understanding the concepts, for the integration languages and makes them compatible [1].

Methods

The terminological sample was taken from BLO. We start defining a kind of acute myeloid leukemia (AML) that has 24 classes. We also intend to define other hematological neoplasms, namely, the Myelodysplastic syndrome with 5 classes and the Myeloproliferative neoplasm with 11 classes. We have a systematized criteria for writing natural language and formal logic language definitions based on the best practices proposed by Köhler et al. [2], Smith et al. [3] and Seppäla and Ruttenberg [4], to mention a few. At this moment, our research does not bring complete literature review, but only few references were adopted for testing the creation of a set of guidelines. The steps in formulating textual definitions are part of our preliminary results, organized in list as follows: a) separate the term; b) get a preliminary definition of the meaning of the term from any source; c) establish the highest genus in the context of use of the term; d) establish the essential characteristic that distinguish the genus from species; e) formulate the first version of the definition in the form S = Def. G which Ds, where "G" stands for genus, which is the parent of S; and "S" stands for species; f) check if the definition is a statement that satisfies the necessary and sufficient conditions; g) check principle of non-circularity; h) check the principle of multiple-inheritance. We don't believe we are contributing to the state of art ontological engineering, but we are making an effort to define some cases in our leukemia ontology. The main contribution of our approach is educational, and an attempt to propose an approach to people building ontologies in the field of Library & Information Science. In general we do not have guidelines to build good definitions [4].

Conclusion

This is an ongoing research project in the Information Science field. We present preliminary and partial results of our research, in the hope that soon we are able to define other types of leukemia.

References

[1] Almeida MB, Proietti ABFC, Ai J, Smith B. The Blood Ontology: an ontology in the domain of hematology. *Proceedings of ICBO*, 2011, US. Available at: http://ceur-ws.org/Vol-833/. Access Apr. 2014.

[2] Köhler J, Munn K, Rüegg A, Skusa A, Smith B. Quality control for terms and definitions in ontologies and taxonomies. *BMC Bioinformatics*. 2006: 7:212.

[3] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies.*Genome Biol* 2005: 6(5):R46.

[4] Seppälä S, Ruttenberg R. Survey on Defining Practices in Ontologies : Report.IN: ICBO 2013, 4, Jul.2013, Montreal, Canada,2013. Available at: http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013_surveyreport_seppalaruttenberg.pdf. Access Nov. 2014.

Address for correspondence

Amanda Damasceno de Souza, Escola de Ciência da Informação, Universidade Federal de Minas Gerais.Rua José Batista Filho, 300, Centro – Sarzedo – MG – Brazil, 32450-000, amanda@ufmg.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1093

A Pilot Study on Modeling of Diagnostic Criteria Using OWL and SWRL

Na Hong[a,b], Guoqian Jiang[a], Jyotishiman Pathak[a], Christopher G Chute[a]

[a] *Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA*
[b] *Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China*

Abstract

The objective of this study is to describe our efforts in a pilot study on modeling diagnostic criteria using a Semantic Web-based approach. We reused the basic framework of the ICD-11 content model and refined it into an operational model in the Web Ontology Language (OWL). The refinement is based on a bottom-up analysis method, in which we analyzed data elements (including value sets) in a collection (n=20) of randomly selected diagnostic criteria. We also performed a case study to formalize rule logic in the diagnostic criteria of metabolic syndrome using the Semantic Web Rule Language (SWRL). The results demonstrated that it is feasible to use OWL and SWRL to formalize the diagnostic criteria knowledge, and to execute the rules through reasoning.

Keywords:

Diagnostic Criteria; Formalization; Rules; OWL; SWRL

Introduction

Standardization and computerization of diagnostic criteria is essential for automating clinical decision making. However, a principal challenge is the lack of an explicit and consistent information model for computable diagnostic criteria. Furthermore, formalizing the inner logic in diagnostic criteria is complex. Semantic Web technologies provide a homogeneous framework that enables an ontology-based modeling using the Web Ontology Language (OWL) and support rules authoring using the Semantic Web Rule Language (SWRL). The objective of this study is to describe our efforts in a pilot study on modeling diagnostic criteria using a Semantic Web-based approach.

Materials and Methods

We first identified a collection of diagnostic criteria from medical textbooks and the literature, and randomly selected 20 diagnostic criteria from different clinical subjects. Second, we manually annotated diagnostic criteria text with the ICD-11 content model [1] and analyzed the distribution of annotations. Third, we invoked the NCBO Bioportal Annotator to analyze the coverage in existing medical terminologies. Fourth, we refined the ICD-11 content model into an operational model using OWL representation. Finally, we performed a case study by manually generating a set of rules in SWRL for the dignostic criteria of metabolic syndrome. We used the Protégé SWRL plugin for rule editing and invoked Pellet as a rule engine to perform reasoning.

Results

Table 1 shows the distribution of data element annotations based on the ICD-11 content model. We found that Investigation Findings and Signs and Symptoms were the two most frequent annotations for data elements in the diagnostic criteria. Terminology coverage analysis using the NCBO BioPortal Annotator demonstrated that SNOMED CT and LOINC have the best coverage for the data element annotations.

Table 1 - Distribution of data element annotations based on ICD-11 Content Model

ICD-11 Content Model	Count	Examples
Investigation Findings	74	Serum triglycerides
Signs and Symptoms	69	Fatigue, Headache
Title	20	Metabolic Syndrome
Causal Properties	18	Pericardial effusion
Classification	12	T71
Severity of Subtype	10	Mind, Moderate, Severe
Body System/Structure	8	Nervous system
Specific Condition	3	Female, Pregnancy
Temporary Properties	2	Age 55, sudden

We constructed SWRL rules for the diagnostic criteria of metabolic syndrome. In total, we produced 21 individual SWRL rules for the criteria. For example, the following is a SWRL rule for the criterion - Waist circumference >102 cm in men. *Rule:Patient(?x), WaistCircum(?y), has_sex(?x, Male), has_result(?x, ?y), has_unit(?y, cm), has_value(?y, ?w), greaterThan(?w, 102) -> has_Evidence(?x, Con1)*. Based on these rules, we invoked Pellet to automatically allocate patients into target diagnosis coded in an ICD category.

Conclusion

In this pilot study, we demonstrated the feasibility of modeling diagnostic criteria using the ICD-11 content model and the terminology coverage of existing biomedical ontologies. We leveraged a Semantic Web-based approach. We believe this study will help future research on the development of formalized representations to capture diagnostic criteria and to enable computer reasoning. We also plan to test the utility of the rules using real-world patient data.

References

[1] ICD-11 Content Model Reference Guide. http://www.who.int/classifications/icd/revision/Content_Model_Reference_Guide.January_2011.pdf?ua=1; Last accessed: December 22, 2014.

Address for correspondence

Dr. Guoqian Jiang, MD, PhD, jiang.guoqian@mayo.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1094

Constructing a Graph Database for Semantic Literature-Based Discovery

Dimitar Hristovski[a], Andrej Kastrin[b], Dejan Dinevski[c], Thomas C. Rindflesch[d]

[a] *Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia*
[b] *Faculty of Information Studies, Novo mesto, Slovenia*
[c] *Faculty of Medicine, University of Maribor, Maribor, Slovenia*
[d] *National Library of Medicine, Bethesda, MD, USA*

Abstract

Literature-based discovery (LBD) generates discoveries, or hypotheses, by combining what is already known in the literature. Potential discoveries have the form of relations between biomedical concepts; for example, a drug may be determined to treat a disease other than the one for which it was intended. LBD views the knowledge in a domain as a network; a set of concepts along with the relations between them. As a starting point, we used SemMedDB, a database of semantic relations between biomedical concepts extracted with SemRep from Medline. SemMedDB is distributed as a MySQL relational database, which has some problems when dealing with network data. We transformed and uploaded SemMedDB into the Neo4j graph database, and implemented the basic LBD discovery algorithms with the Cypher query language. We conclude that storing the data needed for semantic LBD is more natural in a graph database. Also, implementing LBD discovery algorithms is conceptually simpler with a graph query language when compared with standard SQL.

Keywords:

Databases; Data Mining; Semantics; Literature Based Discovery.

Introduction

The goal of LBD is to generate novel hypotheses by analyzing the biomedical literature. For this project, biomedical concepts and relations between them were extracted from MEDLINE and stored as a network. It is common in LBD work to use concept co-occurrence as an indication of a relation between concepts. In this research, we used semantic relations instead of co-occurrences because semantic relations are more expressive and more likely to be real relations. As a source of semantic relations we used the Semantic MEDLINE Database (SemMedDB) which contains semantic relations extracted with the SemRep semantic interpreter (e.g. SemRep extracts the relation Ethionine-CAUSES-Lesion from the sentence "Ethionine can trigger lesions."). SemMedDB has relations from all of MEDLINE and is available as a MySQL database. MySQL is generally efficient, but modeling networks using a relational database causes a large number of many-to-many relations. Complex join queries are then needed to retrieve such data. Here we propose the Neo4j graph database as an alternative for implementing semantic LBD because Neo4j is particularly useful for storing data structured as a network.

Methods

We exported the semantic relation instances from SemMedDB in a text CSV format. We wanted to aggregate the instances into (concept1, relation, concept2, frequency) tuples where "frequency" is the total number of (concept1, relation, concept2) instances. First we tried to aggregate and load the data with the Neo4j "LOAD CSV", but this command was not able to load this amount of data. Then we tried to aggregate with several AWK scripts and to load the data into Neo4j with a stand alone batch importer. This approach worked out very well. Finally, we implemented a few LBD discovery patterns with the Cypher graph query language.

Results

Currently there are 269 047 unique concepts and 14 150 952 distinct relationships between them in the graph database. We show here how to implement the LBD discovery pattern "inhibit the cause of the disease" which can be used to find novel treatments. The idea of the discovery pattern is to find drugs that inhibit some genes that are etiologically related to a disease. Additionally, we are interested only in drugs that have not been already used to treat the disease. In this Cypher query, *phsu*, *gngm* and *dsyn* are UMLS semantic type abbreviations; and *drug*, *gene* and *disease* are variables that are instantiated to particular values when the query is run:

```
MATCH (drug:phsu) -[:INHIBITS]->
(gene:gngm) -[:CAUSES]-> (disease:dsyn)
WHERE NOT (drug)-[:TREATS]->(disease)
RETURN drug, gene, disease
```

Conclusion

Using a graph database such as Neo4j for storing the network data structure needed for semantic LBD is more natural and efficient than using a relational database.

Implementing discovery algorithms in the form of discovery patterns is conceptually easier and simpler when using a graph query language such as Cypher when compared to standard SQL.

Address for correspondence

Dimitar Hristovski e-mail: dimitar.hristovski@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1095

Development of an Ontology to Recommend Exercises from Conceptual Maps

Márcia Ito[a], Débora Lina N. Ciriaco Pereira[a]

[a]*IBM Research, São Paulo, Brazil*

Abstract

The recommendation of exercise plans requires several variables to be considered (e.g., patient's conditions and preferences) and are normally complex to analyze. To facilitate this analysis we proposed the creation of an ontology to assist professionals to recommend exercises. We interviewed 2 experts and this resulted in IDEF diagram and conceptual map. The conceptual map proved to be the preferred way that experts gained more understanding compared with the IDEF diagram. In addition, we also used the conceptual map to validate the formal structure of experts' ideas. From the conceptual map we created an ontology that is being reviewed. After this, we plan to incorporate the ontology into a decision support system that will assist professionals to recommend exercises for their patients.

Keywords:
Biomedical ontologies; Exercise therapy; Medical informatics applications; Clinical decision support systems.

Introduction

The practice of physical exercise has been increasingly encouraged to prevent or aid treatment of chronic diseases. However, in order to have all the benefits, it is necessary for professionals to provide an exercise plan for the specific physical condition of each patient. For this, several variables should be considered. Considering all variables however, makes analysis complex [1]. Thus, we propose to establish a system to support professionals in development of exercise plans which consider both patients' conditions and preferences. Considering these aspects, we decided to develop an ontology to recommend exercises from conceptual maps. This ontology will compose the decision support system.

Materials and Methods

We interviewed 2 experts in the domain area in order to understand their mental models. Following these models, we defined a process of exercise recommendations. Initially, we tried to formalize the process in a flowchart but we realized that creating an IDEF[1] diagram and a conceptual map were better approaches to understand the existing mental models because we could relate concepts by linking words. The results were analyzed by the experts who recommended some adjustments. After this, an ontology was created using the CmapTools[2]. This method is increasingly used once the information has separate sources. Through this method, there results a set of concepts, with their associated relationships and instances which are concise and concrete; and useful for the construction of a decision support system [2].

Results

From the interview with the experts, we created one flowchart, an IDEF diagram and two conceptual maps. An ontology in OWL language was created from the final conceptual map. We formalized the knowledge represented in the conceptual map using an ontology. The tool allowed the creation of a tree that has the structure of the concepts. This tree initially had some inconsistencies and revisions were made in order to improve the ontology.

Discussion and Conclusion

The experts' interviews demonstrated that human knowledge is not structured, and its expression is not a simple task. To structure and express this knowledge, we used several tools. The flowchart proved to be restricted, whereas the IDEF diagram was very technical. On the other hand, the conceptual map proved to be a very useful tool both to structure and validate the mental models of the experts. To build the conceptual map we had to understand the concepts, their connections and interrelations. It was necessary to be careful about linking words so they could eventually give rise to the formal ontology. It was necessary to standardize the terms between both experts; and between the experts and us so we could use the patterns that we had in this area. Even after adjustments in the concept map, the ontology that was generated in a tree format allowed us to recognize problems in the expression of ideas. This demonstrates the need for revisions in the ontology before implementing it in a decision support system.

Acknowledgments

We thank Juliana Batista and Dr. Mauro Vaisberg, the experts who helped us with the plan of exercises.

References

[1] American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins, 2013.

[2] Sarker BK, Wallace P, Gill W. Some observations on mind map and ontology building tools for knowledge management. Ubiquity. 2008;9(9): 1-9.

[1] http://www.idef.com/idef0.htm
[2] http://cmap.ihmc.us/

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1096

Utility of Arden Syntax for Representation of Fuzzy Logic in Clinical Quality Measures

Robert A. Jenders

Center for Biomedical Informatics & Department of Medicine, Charles Drew University, Los Angeles, CA, USA
Department of Medicine, University of California, Los Angeles, CA, USA

Abstract

Background: Prior work has established that fuzzy logic is prevalent in clinical practice guidelines and that Arden Syntax is suitable for representing clinical quality measures (CQMs). Approved since then, Arden Syntax v2.9 (2012) has formal constructs for fuzzy logic even as new formalisms are proposed to represent quality logic. Objectives: Determine the prevalence of fuzzy logic in CQMs and assess the utility of a contemporary version of Arden Syntax for representing them. Methods: Linguistic variables were tabulated in the 329 Assessing Care of the Vulnerable Elderly (ACOVE-3) CQMs, and these logic statements were encoded in Arden Syntax. Results: In a total of 392 CQMs, linguistic variables occurred in 30.6%, and Arden Syntax could be used to represent these formally. Conclusions: Fuzzy logic occurs commonly in CQMs, and Arden Syntax offers particular utility for the representations of these constructs.

Keywords:
Clinical decision support systems; knowledge representation.

Introduction

Fuzzy logic is a multi-valued logic for representing imprecise reasoning. Unlike the typical binary logic employed in knowledge representation (KR) formalisms in clinical medicine, fuzzy logic incorporates degrees of truth or set membership. Clinical quality measures (CQMs) may employ fuzzy logic, using linguistic variables containing modifiers such as “severe” and “partial” without necessarily defining them .

The Arden Syntax standard is a computable formalism for procedural medical knowledge that has been adopted by several vendors. Knowledge is represented in units known as medical logic modules (MLMs). Prior work has established the significant prevalence of fuzzy logic in clinical practice guidelines (100% of a sample had at least 1 linguistic variable) [1] and the utility of the Arden Syntax for representing CQMs, specifically a subset of measures Assessing Care of the Vulnerable Elderly (ACOVE-3) applicable to EHR system and administrative data that were published by RAND in 2007[2]. More recent work has established formal constructs for fuzzy logic representation, including variable declarations and logic statements, in v2.9 of the Arden standard certified in 2012.

Nevertheless, despite this demonstrated prior utility, additional formalisms have been proposed to represent CQMs. The present work was undertaken, in light of recent incorporation of support specifically for fuzzy logic representation in contemporary versions of Arden, to assess the value of these constructs and the consequently improved utility of Arden for representation of CQMs.

Methods

Expanding from the subset of 36 ACOVE-3 CQMs previously used in the assessment of the utility of Arden Syntax as a formalism for CQMs, all 392 ACOVE-3 CQMs were tabulated for the presence of linguistic variables. Logic statements containing these variables then were encoded in Arden Syntax.

Results

The instrument (VES-13) used to determine that an elder is vulnerable, which is a condition of all the ACOVE-3 CQMs, itself contains 3 self-evaluation linguistic variables applied to 6 different criteria. Of the 392 CQMs, 120 (30.6%) contained at least 1 linguistic variable. Four of these contained 2 linguistic variables to yield a total of 124 in ACOVE-3.

Conclusions

A large corpus of CQMs had fewer measures (30.6%) with linguistic variables than a prior corpus of clinical practice guidelines (100%). Nevertheless, fuzzy logic is sufficiently prevalent in CQMs such that, when coupled with prior evidence of its utility for representing CQMs, Arden Syntax--with its new formal constructs for fuzzy logic--offers particular utility for such representation.

Acknowledgements

This work was supported by NIMHD grant U54MD007598 (“AXIS”) and NCATS grant UL1TR000124 from the National Institutes of Health (USA).

References

[1] Jenders RA. Potential utility of fuzzy logic constructs in the Arden Syntax. AMIA Annu Symp Proc. 2009:894.

[2] Jenders RA. Suitability of the Arden Syntax for representation of quality indicators. AMIA Annu Symp Proc. 2008; 991.

Address for correspondence

Robert A. Jenders, MD, MS, FACP, FACMI
Center for Biomedical Informatics, Charles Drew University
Department of Medicine, University of California, Los Angeles
1731 E 120th St, Building N, Los Angeles, CA 90059 USA
jenders@ucla.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1097

Developing a Standards-Based Information Model for Representing Computable Diagnostic Criteria: A Feasibility Study of the NQF Quality Data Model

Guoqian Jiang[a], Harold R. Solbrig[a], Jyotishman Pathak[a], Christopher G. Chute[a]

[a] *Mayo Clinic College of Medicine, Rochester, MN, USA*

Abstract

The lack of a standards-based information model has been recognized as a major barrier for representing computable diagnostic criteria. In this paper we describe our efforts in examining the feasibility of the Quality Data Model (QDM)–developed by the National Quality Forum (NQF)–for representing computable diagnostic criteria. We collected the diagnostic criteria for a number of diseases and disorders (n=12) from textbooks and profiled the data elements of the criteria using the QDM data elements. We identified a number of common patterns informed by the QDM. In conclusion, the common patterns informed by the QDM are useful and feasible in building a standards-based information model for computable diagnostic criteria.

Keywords:

Diagnostic Criteria; Quality Data Model (QDM); ICD-11 Content Model; ISO 11179 Model.

Introduction

For decades clinicians have recognized that their practices greatly benefit from scientifically-based diagnostic criteria, iteratively improved through ongoing research [1]. With recent advances in computerized patient records systems there is an urgent need for producing computable and standards-based diagnostic criteria that could be utilized more effectively in clinical and translational applications, ultimately improving patient care. The lack of a standards-based information model has been recognized as a major barrier for achieving computable diagnostic criteria. The objective of this study is to describe our efforts in developing a standards-based information model for computable diagnostic criteria.

Methods

We established a library of diagnostic criteria for a collection of diseases and disorders. We profiled these diagnostic criteria using QDM elements. National Quality Forum (NQF) developed the Quality Data Model (QDM), which is an information model for representing electronic health record-based quality "eMeasures" [2]. We identified individual criteria from each disease or disorder and profiled them with QDM elements. We analyzed the distribution of QDM elements and identified common patterns for representing the diagnostic criteria. We developed an information model for representing diagnostic criteria by integrating QDM and ICD-11 content models in a Semantic Web-based framework.

Results

We searched electronic medical textbooks from the Mayo Clinic Library website using keywords "Diagnostic Criteria" and selected a collection of diagnostic criteria from 12 diseases or disorders. In total, 139 individual criteria were identified and used as a test set for further analysis. We profiled all 139 criteria using QDM elements in terms of QDM categories, datatypes, and attributes. In total we identified eight QDM datatypes and their attributes from eight distinct QDM categories; of these QDM datatypes, five are most commnly used: "Laboratory Test, Performed;" "Diagnostic Study, Performed;" "Diagnosis, Active;" "Physical Exam, Performed;" and "Symptom, Active."

Discussion

The study is motivated by the requirement of the WHO ICD-11 revision project, in which the creation of diagnostic criteria for the diseases and disorders in ICD-11 has been proposed as a key component for its revision. In this study we used a data-driven approach to evaluate the feasiblity of using QDM to represent diagnostic criteria. We developed a standards-based information model informed by the QDM and ICD-11 content models. We envision that the QDM-based information model and tools for representing quality measures could be used to tackle the challenges in achieving computable diagnostic criteria. This study suggests that the common patterns informed by QDM are useful and feasible in building a standards-based information model for computable diagnostic criteria.

Acknowledgments

This work is supported in part by funding from the caCDE-QA (1U01CA180940-01A1), PhEMA (R01 GM105688), and a Mayo-WHO Contract 200822195-1.

References

[1] Yager J, McIntyre JS. DSM-5 Clinical and Public Health Committee: challenges and considerations. The American Journal of Psychiatry. 2014 Feb;171(2):142-4.

[2] Quality Data Model. 2014 [cited December 15, 2014]; Available from: http://www.healthit.gov/quality-data-model

Address for correspondence

Correspondence to: Dr. Guoqian Jiang, MD, PhD at jiang.guoqian@mayo.edu.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1098

A Standards-based Semantic Metadata Repository to Support EHR-driven Phenotype Authoring and Execution

Guoqian Jiang[a], Harold R. Solbrig[a] , Richard Kiefer[a], Luke V. Rasmussen[b], Huan Mo[c], Peter Speltz[c], William K. Thompson[b], Joshua C. Denny[c], Christopher G. Chute[a], Jyotishman Pathak[a]

[a] Mayo Clinic College of Medicine, Rochester, MN, USA;
[b] Northwestern University, Chicago, IL, USA;
[c] Vanderbilt University, Nashville, TN, USA.

Abstract

This study describes our efforts in developing a standards-based semantic metadata repository for supporting electronic health record (EHR)-driven phenotype authoring and execution. Our system comprises three layers: 1) a semantic data element repository layer; 2) a semantic services layer; and 3) a phenotype application layer. In a prototype implementation, we developed the repository and services through integrating the data elements from both Quality Data Model (QDM) and HL7 Fast Healthcare Inteoroperability Resources (FHIR) models. We discuss the modeling challenges and the potential of our system to support EHR phenotype authoring and execution applications.

Keywords:

Metadata Repository; Clinical Phenotyping; Quality Data Model (QDM); HL7 FHIR.

Introduction

The Quality Data Model (QDM) is an information model developed by the National Quality Forum (NQF) and a promising candidate for representing EHR-driven phenotyping algorithms for clinical research [1]. In this study, we extend the Semantic Web-based framework of a previous study that provides a standards-based, semantically annotated, machine-readable rendering of the QDM [2], and develop a semantic metadata repository and associated Web services by integrating HL7 FHIR data element models [3]. Integrating the data elements provides a more comprehensive coverage for clinical phenotype applications.

Methods

Our system is comprised of first a semantic data element repository layer, in which we leverage both W3C standards, such as Resource Description Framework (RDF) and Web Ontology Language (OWL), and the meta-data standard ISO 11179 [4] to describe the QDM reference model, data model elements and logic elements. The second layer is a semantic services layer, while the third layer is a phenotype application layer.

In our previous study [2], we developed a QDM schema in OWL representing the QDM reference model. In this study, we extended the schema with the notions of HL7 FHIR Datatypes and Resources, and is designed as a natural extension of the ISO 11179 standard. We populated the schema with data elements from HL7 FHIR models as QDM schema instances (Table 1), and developed RESTful services on the repository (https://github.com/PheMA/phema-mdr), being utilized by a phenotype authoring tool under active development.

Table 1 – Populated data elements

	QDM	HL7 FHIR	Examples (FHIR)
Category	18	99	Medication
Datatype	76	99	Medication
Attribute	528	1021	Medication Kind
Value Set	-	180	Medication Kind
Logic Element	53	-	-

Conclusion

Our system provides a standards-based semantic infrastructure in enabling data element services to support phenotype authoring and execution. In future work, we plan to develop a standard interface mechanism with Clinical Information Modeling Initiative (CIMI)-compliant clinical models.

Acknowledgments

This work has been supported by funding from PhEMA (R01 GM105688), eMERGE (U01 HG006379, U01 HG006378 and U01 HG006388), and caCDE-QA (1U01CA180940-01A1).

References

[1] Thompson WK, Rasmussen LV, Pacheco JA, Peissig PL, Denny JC, Kho AN, Miller A, Pathak J. An evaluation of the NQF Quality Data Model for representing Electronic Health Record driven phenotyping algorithms.AMIA Annu Symp Proc. 2012;2012:911-20.

[2] Jiang G, Solbrig HR, Kiefer R, Rasmussen LV, Mo H, Speltz P, Thompson WK, Denny JC, Chute CG, Pathak J. A semantic framework for quality data model to support EHR-driven phenotype authoring and execution. AMIA Jt Summits Transl Sci Proc. 2015 (in press).

[3] HL7 FHIR. http://www.hl7.org/implement/standards/fhir/; Last visited at December 1, 2014

[4] ISO/IEC 11179, Information Technology -- Metadata registries (MDR). http://metadata-standards.org/11179/; Last visited at December 1, 2014.

Address for correspondence

Correpondence to Dr. Guoqian Jiang at jiang.guoqian@mayo.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1099

Development of an Adolescent Depression Ontology for Analyzing Social Data

Hyesil Jung, RN[a], Hyeoun-Ae Park, PhD, RN, FAAN[a,b], Tae-Min Song, PhD[c], Eunjoo Jeon, MS, RN[a], Ae Ran Kim, MS, RN[a], Joo Yun Lee, MS, RN[a]

[a] College of Nursing, Seoul National University, [b]Research Institute of Nursing Science, Seoul National University, Seoul, Korea, [c]Korea Institute for Health and Social Affairs

Abstract

Depression in adolescence is associated with significant suicidality. Therefore, it is important to detect the risk for depression and provide timely care to adolescents. This study aims to develop an ontology for collecting and analyzing social media data about adolescent depression. This ontology was developed using the 'ontology development 101'. The important terms were extracted from several clinical practice guidelines and postings on Social Network Service. We extracted 777 terms, which were categorized into 'risk factors', 'sign and symptoms', 'screening', 'diagnosis', 'treatment', and 'prevention'. An ontology developed in this study can be used as a framework to understand adolescent depression using unstructured data from social media.

Keywords:

Depression, Adolescent, Ontology, Social media

Introduction

According to OECD Health Data, the suicide rate in South Korea was 29.1 per 100,000 people in 2012, the highest among OECD members states. Moreover, suicide is the leading cause of death among South Korean adolescents. According to a study, approximately 80% of suicides are related to depression. Adolescent depression has the course of chronic relapsing and affects interpersonal realationships and learning ability. Therefore, early detection and intervention are important considerations. Nowadays, big data analytical techniques and extraction methods to derive valuable information at the individual level, are in the spotlight. In Korea, 77.1% of youth own a SNS account[1] and they record their daily activities and openly express their emotions on their social network. Therefore, it is possible for us to use social analytics to monitor activities and sentiments of these adolescents. The aim of this study is to develop an ontology that can be used as a framework for analyzing SNS data to detect danger of adolescent depression and provide timely care.

Methods

An ontology was developed based on 'Ontology development 101'[2]. First, we reviewed exsisting ontologies and then collected important terms in the ontology from several clinical practice guidelines and Twitter postings and enumerated them. Second, we defined the classes and the class hierarchy of the extracted terms. Third, we defined the properties of classes and created instances. Finally, we developed an ontology linking these classes.

Results

We searched on Pubmed and Google Scholar using the keywords 'adolescents'/youth depression ontology'. However, there were no ontologies available. We extracted 777 terms from clinical practice guidelines and Twitter. In this study, six domains(top-classes) and 134 sub-classes were drawn from the enumerated terms. 'Risk Factors' range from individual, family, and social context such as school and community. After 'screening' the 'risk factors' and 'sign & symptoms' of individual, we made a 'diagnosis' of depression. If an adolescent is diagnosed with depression, 'treatment' will be needed. However, if he/she is determined to be a normal or vulnerable individual, 'prevention' services will be followed. Figure 1 illustrates relationships among classes in ontology.

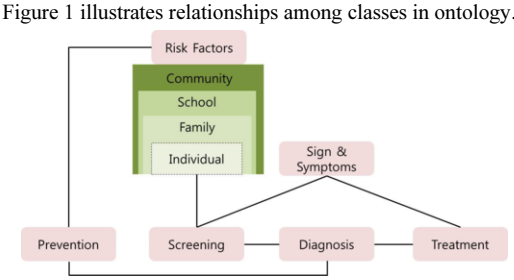


Figure 1. An ontology of adolescent depression

Conclusion

We have developed an adolescent depression ontology based on terms extracted from the clinical practice guidelines and SNS. The ontoloy developed in this study will be used as a framework to monitor youth depression using social media data.

References

[1] Ministry of Gender Equity & Family. 2013 Study of media use among children and youth. 2013.

[2] Noy, N. McGuinness, D. Ontology Development 101: A Guide to Creating Your First Ontology. Standford Knowledge Systems Laboratory Technical Report. 2001

Acknowledgements

This Work was supported by the ICT R&D Program of MSIP[I0114-14-1016] and the Creative Vitamin Project.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP)(No.20210-0028631

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1100

A Framework for Modeling Workflow Execution by an Interdisciplinary Healthcare Team*

Mounira Kezadri-Hamiaz[a], Daniela Rosu[a], Szymon Wilk[b], Craig Kuziemsky[a], Wojtek Michalowski[a], Marc Carrier[c]

[a] *Telfer School of Management, University of Ottawa, Canada*
[b] *Institute of Computing Science, Poznan University of Technology, Poland*
[c] *Ottawa Hospital Research Institute, The Ottawa Hospital, Canada*

Abstract

The use of business workflow models in healthcare is limited because of insufficient capture of complexities associated with behavior of interdisciplinary healthcare teams that execute healthcare workflows. In this paper we present a novel framework that builds on the well-founded business workflow model formalism and related infrastructures and introduces a formal semantic layer that describes selected aspects of team dynamics and supports their real-time operationalization.

Keywords:
Patient Care Team; Workflow; Health Information Systems.

Introduction

Patient management processes are being standardized and formalized as workflows, which in healthcare environment are executed by an interdisciplinary healthcare teams (IHTs). While there has been extensive research on the formalization and execution of business workflows [1], its applicability in the healthcare domain is limited. Limited expressiveness of the business workflow models restricts the ability to represent operations of an IHT, in particular the role of Most Responsible Physician (*MRP)*, team member's role variability and membership dynamics [2]. Role variablity implies that team members may play multiple and different roles when executing workflow tasks, while membership dynamics means that members may join and leave team at any time. Thus, to improve expressiveness of workflow models, a dedicated modeling framework needs to formalize the structure and behavior of an IHT.

Methods

In this poster, we present the *Team and Workflow Management Framework* (TWMF) that extends business workflow models to support the behavior (understood as team formation, management and task-practitioner allocations) of an IHT. The TWMF is designed to be executed in an environment where multiple patients are managed simultaneously by different IHTs and it relies on a set of assumptions that are briefly discussed below. An IHT is formed when a new patient with a certain presentation is admitted and her management (according to a disease-specific workflow) starts. Each IHT includes a MRP who is responsible for patient management and has special duties such as handling exceptional situations. The IHT members are recruited and released dynamically from among available practitioners, as dictated by the execution of the tasks in the workflow. The recruitment and assignment of team members to tasks is done by matching their clinical expertise with the expertise required for completing the tasks. We assume that a practitioner possesses several regularly updated capabilities each with a specific competency level, and that a task execution requires one or more capabilities with associated competency levels. A practitioner can be assigned to a task if she possesses the required capability at a competency level that is at least as good as the competency level identified for the task. TWMF introduces a semantic layer that fomalizes all these aspects through the ontology describing all required concepts, behavioral rules describing team behavior, and an instance base containing all necessary instances of the ontology concepts. The semantic layer is encoded in First-order logic (FOL) [3] formalism.

Results

TWMF is operationalized through a *Team and Workflow Controller* which interfaces with the *Hospital Information System* and the FOL *Reasoner* that operates on the semantic layer and drives workflow execution by an execution engine (such as IBM's Business Process Manager).

Conclusions

We have presented a novel framework for modeling workflow execution by an IHT. It relies on a semantic layer that captures team behavior and team dynamics.

We are testing a framework for different clinical scenarios and we are working on its extension to include durative tasks.

References

[1] van Der Aalst WM, et al. Business process management: a survey. Business Process Management: Springer; 2003: 1-12.
[2]Andreatta PB. A typology for health care teams. Health Care Management Review. 2010;35(4): 345-54.
[3]Smullyan RM. First-order logic: Springer; 1968.

* This research was supported by grants from the NSERC CREATE Program and the Telfer School of Management.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1101

Evaluating a Hierarchical Clinical Event Linkage Model for Clinic-Specific Databases

Justin Liu, BASc[a], Tran Truong, BMath[a]

[a] *Health Informatics Research, University Health Network, Toronto, ON Canada*

Abstract

A relational database model is presented that stores the hierarchical linkages between clinical events with qualifier codes, such that the explicit contextual meaning of an event's attributes is preserved upon retrieval. A retrospective analysis of 302 forms built upon the model showed that 91% of 17,899 data elements requested by clinicians and researchers from 19 clinics were successfully represented, but that 62% were never used more than once. These results reinforce the specificity of clinic-specific databases and the need for unambiguous, explicitly-stored clinical data.

Keywords:

Databases, Factual; Patient-Specific Modeling; Models, Computer; Models, Decision Support; Medical Records.

Introduction

The Explicit Clinical Event Linkage (ExCEL) data model was created to adapt to the granularity and variation of data requirements inherent in different clinical and research settings; this domain-specific data is not typically captured in standard EHR's or contained in the narrative of clinical notes [1,2]. The model is currently implemented in relational databases – such as Oracle, MySQL, and MSSQL – for an array of software (eCancerCare, CKCIS, eColposcopy, UHN OR Notes) to store clinical data for 302 forms for 19 clinics across 5 institutions. The goal of the ExCEL model is to avoid misinterpretations and medico-legal issues that arise from data ambiguity and duplication when retrieving data for clinical documentation, research studies, and decision support.

Methods

The ExCEL model captures 36 types of clinical event entities. Each clinical entity stores atomic, single-valued attributes with distinct clinical meaning (e.g. a procedure entity has a start date attribute); eligible attributes are coded using SNOMED CT. However, a biopsy procedure date can be captured in a clinic note for a past biopsy in the patient's history, for a biopsy performed at the visit, or for a planned biopsy. In order to also store the contextual meaning of the procedure date attribute, hierarchical linkages between clinical events are stored with qualifier codes (SNOMED CT where possible) with explicit meanings. Data fields requested by clinician stakeholders led to very complex hierarchies of events, e.g., "the interventional outcome of a bipolar coagulation performed as an intervention for esophagitis observed in an esophageal exam performed for a gastroscopy as part of an endoscopy operative report". Any data elements that could not be modeled were handled using a less efficient Entity-Attribute-Value design [3]. A retrospective analysis of all data elements was performed to determine the elements that were successfully represented in the ExCEL model, and the non-recurring elements that were used only once.

Results

Table 1 displays the percentages of modeled and non-recurring data elements classified by type of form. Non-modelable elements were specific to granular clinical situations, e.g., "# of children had before removal of testicles" or "# of 8mm suprapubic ports placed in preparation for prostatectomy".

Table 1– Percentages of Non-Recurring and Missing Elements

Type of form	# elements	% modeled	% nonrecur.
All forms	17899	91%	62%
Surg. Reports	5485	76%	83%
Clinic notes	9437	97%	58%
Other forms	2977	98%	23%

Conclusion

The ExCEL model successfully stores data for 91% of 17,899 data elements requested by clinicians, but 62% are never used more than once. The higher percentage of non-recurring data elements in surgical reports and clinic notes is indicative of the large amount of variation between procedures and disease treatment workflows, which demonstrates the specificity of clinic-specific databases and reinforces the difficulty of creating a standardized clinical data model. This specificity must be preserved, however, as extending the meaning of data beyond what its data model is designed to describe can involve making incorrect assumptions with medico-legal implications when interpreting the data upon retrieval.

References

[1] Shenvi EC, Meeker D, Boxwala AA. Understanding data requirements of retrospective studies. Int J Med Inform. 2015;84(1):76-84. Epub 2014 Oct 12.

[2] Johnson S, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P. An Electronic Health Record Based on Structured Narrative. Am Med Inform Assoc. 2008; 15(1): 54-64.

[3] Chen R, Nadkami P, Marenco L, Levin F, Erdos J, Miller P. Exploring Performance Issues for a Clinical Database Organized Using an Entity-Attribute-Value Representation. Am Med Inform Assoc. 2000; 7(5): 475-487.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1102

Enabling Self-Monitoring Data Exchange in Participatory Medicine

Guillermo Lopez-Campos[a], Bahadorreza Ofoghi[b], Fernando Martin-Sanchez[a]

[a] *Health and Biomedical Informatics Centre, The University of Melbourne, Melbourne, VIC, Australia*
[b] *Department of Computer and Information Systems, The University of Melbourne, Melbourne, VIC, Australia*

Abstract

The development of new methods, devices and apps for self-monitoring have enabled the extension of the application of these approaches for consumer health and research purposes. The increase in the number and variety of devices has generated a complex scenario where reporting guidelines and data exchange formats will be needed to ensure the quality of the information and the reproducibility of results of the experiments. Based on the Minimal Information for Self Monitoring Experiments (MISME) reporting guideline we have developed an XML format (MISME-ML) to facilitate data exchange for self monitoring experiments. We have also developed a sample instance to illustrate the concept and a Java MISME-ML validation tool. The implementation and adoption of these tools should contribute to the consolidation of a set of methods that ensure the reproducibility of self monitoring experiments for research purposes.

Keywords:

Self-monitoring; Data Exchange; Standards; Information Sharing.

Introduction

In recent years we have witnessed an increase in the number of new devices and apps developed for health purposes and self-monitoring. These systems are also being used for biomedical research purposes. Despite of the increase in their use, there is still a great disparity in both the data formats, data analysis and data management approaches as well as on the metadata required for their interpretation. Based on previous successful experiences in other disciplines, we have recently developed a minimal information reporting guideline for self monitoring experiments (MISME) [1], this guideline defined the minimal metadata required for the interpretation of these experiments.

Once a minimal reporting guideline has been proposed, the next step is the development of information processing tools based on those reporting guidelines. Facilitating data exchange in a standardised format represents a crucial step in the development of minimal reporting guidelines. For this purpose, in other disciplines (such as bioinformatics) extended markup languages (XML) have been developed.

In this work we present an implementation of an XML-based data exchange format for self monitoring experiments (MISME-ML) based on MISME as well as a MISME-ML validation tool.

Methods

For the development of the data exchange format we have chosen an extensible mark-up language which captures the information included in the MISME reporting guideline. To facilitate the design and creation of this XML format, we transformed the concepts covered in the MISME reporting guideline into an entity-relationship model. An XML format validation tool was also developed using JAVA.

Results

We have modeled and developed an XML format (MISME-ML) to facilitate the exchange of self monitoring data according to our previously developed minimal reporting guideline MISME. Together with this XML format we have developed a MISME-ML validator tool to assess the quality of the MISME-ML files, checking the different elements that are part of MISME and the nature of the relationships among them, ensuring the existence of the correct relationships among the different elements and avoiding having unrelated data.

Conclusions

With the expansion of self monitoring approaches for research purposes and health applications, there is an increasing need of using standards and methods to ensure the reproducibility and sharability of the results from these experiments.

The development of a data exchange format and the companion XML validation tool come to provide and extend the functionalities related with the minimal information reporting guideline aiming to facilitate the process of data exchange and the reproducibility of the experiments.

Acknowledgments

This work has been supported by an ECR grant of the University of Melbourne.

References

[1] Lopez-Campos G, Almalki M, Martin-Sanchez F. Proposal for a Standardised Reporting Guideline to Annotate Health-related Self-Quantification Experiments. Stud Health Technol Inform. 2014;202:79-82

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1103

DServO: A Peer-to-Peer-based Approach to Biomedical Ontology Repositories

Zakaria Mambone[a], Mahamadi Savadogo[a], Borlli Michel Jonas Some[a], Gayo Diallo[b]

[a] *Polytechnic University of Bobo-Dioulasso, Computer Sciences Institute, Burkina Faso*
[b] *University of Bordeaux, INSERM U897, F-33000, Bordeaux, France*

Abstract

We present in this poster an extension of the ServO ontology server system, which adopts a decentralized Peer-To-Peer approach for managing multiple heterogeneous knowledge organization systems. It relies on the use of the JXTA protocol coupled with information retrieval techniques to provide a decentralized infrastructure for managing multiples instances of Ontology Repositories.

Keywords:
Ontology Repository, Biomedical Ontologies, Peer-To-Peer.

Introduction

Presently, new Knowledge Organization Systems (KOS) are proliferating rapidly, ranging from terminologies and structured vocabuaries to formal ontologies with well-defined semantics. As a consequence, there is a need of systems and tools for managing these KOS's and retrieving them in an efficient way in order to be reusable in knowledge-based applications. Several systems have been proposed to address this issue. BioPortal [1] is presently the most popular of these within the biomedical domain. Most of the systems follow a client-server model. This paradigm results in a number of limitations. First, it is not fault tolerant, since a server failure results in data unavailability. Second, when a user publishes an ontology onto a server, they lose the ability to alter it.

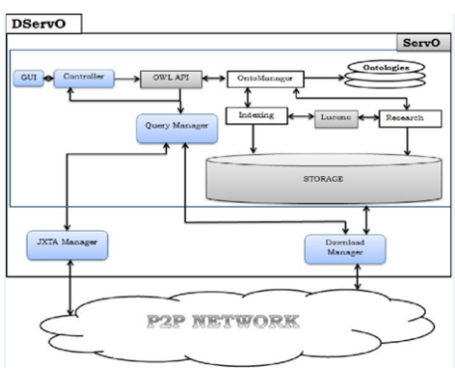


Figure 1 – Architechture of a DServO Peer

In this poster, we propose DServO, an extension of the ServO ontology server [2], which offers a decentralized approach – based on a Peer-To-Peer (P2P) paradigm – for managing multiple ontology servers and making them interoperable.

Methods

We have extended ServO by adding a P2P layer. To do so, we relied on the JXTA technology[1] to create a P2P network of mutiple extended ServO (DServO) instances. Each instance of DServo is constitued by the following components (Figure 1): i) A graphical user interface (GUI), which allows sending queries and retrieving responses; ii) a controller, which takes charge of dispatching user queries to the query manager, and offers services for managing ontologies; iii) a query manager, which mediates between the GUI and controller, and the network layer (JXTA manager); iv) a JXTA manager, which takes charge of all connection management and of discovering peers within the P2P network; and v). a retrieval module which takes charge of retrieving ontologies and ontological entities from the P2P network.

Results

We have implemented DServO in Java using the OWL API as an ontology manager, and the Apache Lucene as an information retrieval library. The system is able to handle efficiently very large OWL and RDF KOS's with hundreds of thousands of entities – such as SNOMED CT and the NCI Thesaurus – in a timely manner.

Conclusion

DServO has been succefully evaluated by instatiating 4 nodes of ontology servers and handling various – sometimes duplicated – biomedical KOS's.

References

[1] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 2009, 37(Web Server):W170–W173.

[2] Diallo G: Efficient Building of Local Repository of Distributed Ontologies. In: IEEE Proceeding of SITIS'2011. Dijon, France, 2011:159–166.

Address for correspondence

Gayo.Diallo@u-bordeaux.fr

[1] https://jxta.kenai.com/

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1104

OntoMama: An Ontology Applied to Breast Cancer

M.T.D. Melo[a], V.H.L. Gonçalves[a], H.D.R Costa[a], D.S Braga[a], L.B Gomide[a], C.S Alves[a], L.M. Brasil[a]

[a] *University of Brasília (UnB) at Gama (FGA), Brazil*

Abstract

This article describes the process of building an ontology to assist medical students and professionals specialized in Oncology. The ontology allows the user to obtain knowledge more quickly and thus assist professionals in their decision-making.

Keywords:
Breast Cancer, Ontology, OntoMama.

Introduction

Technological advances are having a strong impact in the health sector. Information technology, for example, which seeks improvements in the processing and dissemination of data, assist health professionals in their work [1].

The human body consists of several structures organized hierarchically. For the process of learning the human body, including the study of dysfunctions, it is necessary to organize these structures according to their concept, relationships and functions, in order to facilitate identi-fication to these elements [2].

An example of pathology that affects the functioning of the organism is breast cancer. This type of cancer develops in the mammary gland as a result of genetic changes in some set of breast cells [2]. This article presents the development of OntoMama, an ontology to assist professionals and students specialized in Oncology. It consists of a query web service that provides conceptual information, images, videos and prototypes developed in an environment of three-dimensional (3D) modeling.

Materials and Methods

The domain and scope established for the development of the ontology are linked to breast cancer. The methodology adopted was the METHONTOLOGY and the construction tool chosen was the editor for ontologies and knowledge bases Protégé.

Results

Figure 1 shows the topics that compose the ontological model. This model is partially implemented because there is a need to standardize the structures with bibliographic material.

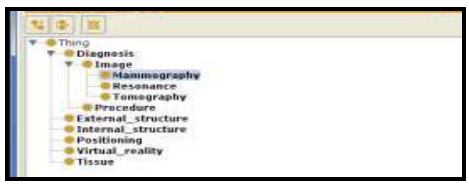


Figure 1 – OntoMama Model.

Conclusion

This article introduced the development of an ontology for to the study of breast cancer. The next step is to complete the ontology with textual content, 3D models, images, and videos for each element. The implementation phase will be finalized with an Intelligent Tutor System (ITS) based on the ontological model developed [3].

Acknowledgements

This work is being funded through the public notice Nº.15/2008CT/CNPq/FNDCT/CAPES/FAPEMIG/FAERJ/FAPES, National Institutes of Science and Technology, and for the support of INCT/MACC. The first author is grateful for the financial support provided by the CNPq to IC research grant.

References

[1] Baracho RMA, Magalhães AMD. Information management as support for reducing uncertainty in decision making: case study of the Instituto Federal de Minas Gerais-IFMG. Information & Technology, 2014; 1(1):81-95.

[2] BREAST CANCER. Available from: http://www.cancer.org/acs/groups/cid/documents/webcontent/003090-pdf.pdf.

[3] Melo JSS. Service oriented architecture for integration of technologies applied to a Three-dimensional Breast Anatomy interactive Atlas. PhD thesis in electrical engineering. Brasília: University of Brasília, 2012.

Address for correspondence

Maria Tereza Dourado Melo
terezadourado11@gmail.com
Lourdes Mattos Brasil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1105

National Healthcare Policies in Chile: An Ontological Meta-Analysis

Alicia Núñez Mondaca[a], Arkalgud Ramaprasad[b], Thant Syn[c]

[a]Universidad de Chile, [b]University of Illinois at Chicago, [c]Texas A&M International University

Abstract

We present an ontological meta-analysis of the national healthcare policies in Chile. Using a logically constructed ontology based on the common body of knowledge as a lens, we map the 39 key policies. The ontological map provides a synoptic, systematic, and systemic view of the policies, and highlight their emphases and biases.

Keywords:

Health Care Economics and Organizations [N03]; Health Care Quality, Access, and Evaluation [N05], Health Services Administration [N04], Health Policy [I01.655.500.608.400]

Introduction

The evolution of the laws indicates that Chile is moving towards universal access to affordable and quality healthcare. However, it is necessary at this point to have a global view of the policies that have been implemented in the country, assess them, and then decide the future direction of the healthcare system. Ontological meta-analysis will help us to efficiently formalize, standardize, and manage the information regarding the policies that have shaped today´s healthcare system, and design future policies to fill the gaps.

Results

The ontological map of monads – individual categories in the ontology – is shown in Figure 1. The number in parentheses adjacent to the category indicates its frequency of occurrence in the 39 policies. The bar below the category is a visual indicator of the same scaled to the maximum number of occurrences of any one category (National – 39). On the one hand, Financial and Regulatory Policies and National focus are highly emphasized; Technology policies and Satisfaction and Timeliness outcomes are hardly emphasized. Yet, almost all elements in the ontology find expression in at least one policy – an indicator of their extensive coverage.

Discussion and Conclusion

A majority of the policies are financial, regulatory, and administrative; just a few of them target the insurance structure. The country needs policies intended to create a system for financing health services that at the same time is adequately managed. A system that guarantees outcomes such as quality and safety, which have been addressed in past health policies, should always be a priority. Policies should also provide access to essential medicine and technologies

References are available on request

Address for correspondence

Alicia Núñez Mondaca; E-mail: anunez@fen.uchile.cl; Phone: +5622978344

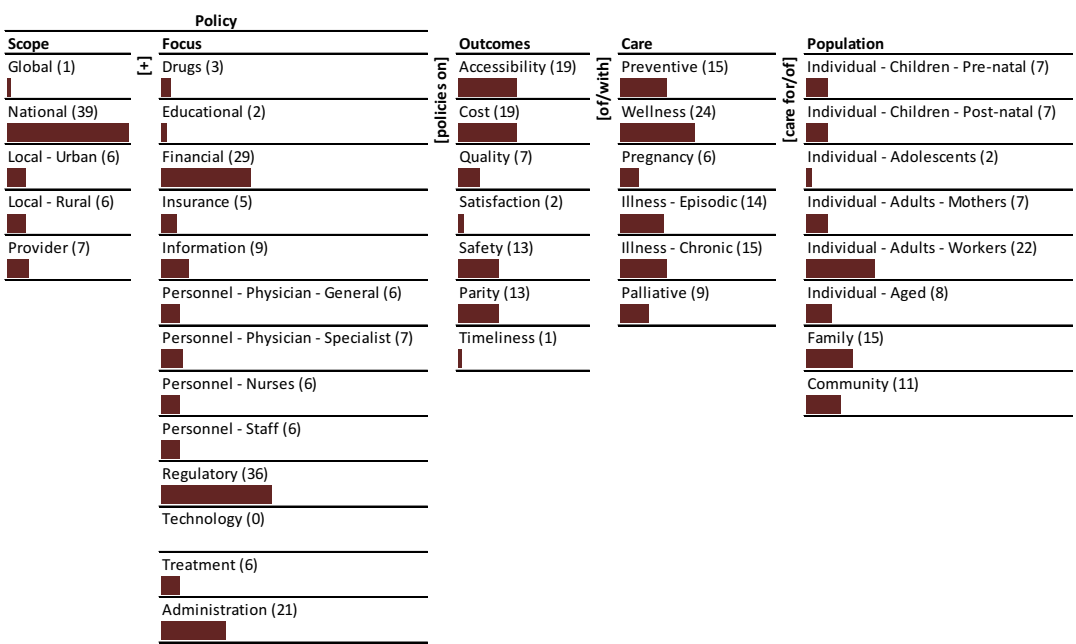


Figure 1: Ontological Map of National Healthcare Policy in Chile (Monads)

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1106

Modelling the Medication Management System for Resource Limited Settings: A Formal Representation of the Prescribing and Dispensing Phases

William Ogallo, Andrew S. Kanter

Department of Biomedical Informatics, Columbia University, New York, USA

Abstract

We propose a conceptual data model for relational databases targeting the prescribing and dispensing phases of the medication management system. The model was developed using recommendations from existing standards and guidelines, with necessary modifications made to suit adoption in resource-limited settings. We present the model as an entity-relationship diagram with 10 entities, 12 relationships and 48 attributes. It is our hope that this work will help mitigate barriers in the implementation of electronic prescribing and dispensing standards in the developing world.

Keywords:

Medication Systems, Hospital; Developing Countries.

Introduction

The medication management system refers to the continuum of prescription writing, prescription transmission, medication dispensing, medication administration and medication therapy monitoring. While standards and recommendations for modelling the electronic information within this domain are available, they primarily exist in the form of textual narratives. The goal of this study was to create and visualize a conceptual model for the prescribing and dispensing phases of the medication management system primarily targeting implementation in resource-limited settings.

Methods

We scrutinized existing literature evidence to understand the workflows, processes, activities, tasks, actors and data elements associated with the two phases. We synthesized the information to understand what data needed to be stored and how these should be structured. We then developed a high-level semantic description of the data and their associated constraints, using the Entity-Relationship model.

Results

The model has a total of 10 principal entities, 12 one-to-many relationships and 48 attributes (Figure 1). The key actors are the *Patient*, *Prescriber* and *Dispensing Staff* who must be uniquely identified. The *Prescription* entity refers to the order for supply of a medication to a particular patient. Attributes of the *Prescribed Drug Information* entity are the action information about the prescribed drug. The *Dispensing Record* entity is the record in response to a prescription showing that a particular medication was supplied to a particular patient and instructions of administration provided. The *Dispensed Medication Information* entity contains attributes that describe the dispensed medication and the *Refill Record* entity is the record tracking the refill history of every dispensed medication.

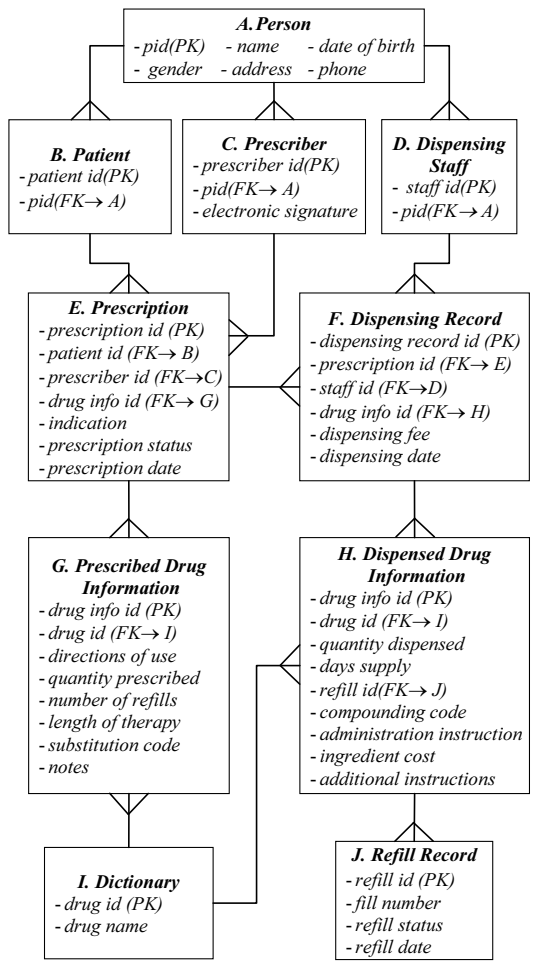


Figure 1 - Entity-Relationship Diagram for the Prescribing and Dispensing Phases

Conclusions

We propose a conceptual model for electronic prescribing and dispensing databases. The model is consistent with current standards and is applicable in resource-limited settings.

Address for correspondence

William Ogallo woo2000@cumc.columbia.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1107

Real-time Data Fusion Platforms: The Need of Multi-dimensional Data-driven Research in Biomedical Informatics

Satyajeet Raje[a], Bobbie Kite[b], Jay Ramanathan[a], Philip Payne[b]

[a] *Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA*
[b] *Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA*

Abstract

Systems designed to expedite data preprocessing tasks such as data discovery, interpretation, and integration that are required before data analysis drastically impact the pace of biomedical informatics research. Current commercial interactive and real-time data integration tools are designed for large-scale business analytics requirements. In this paper we identify the need for end-to-end data fusion platforms from the researcher's perspective, supporting ad-hoc data interpretation and integration.

Keywords:
Research informatics; knowledge bases; data curation.

Introduction

Data-driven research and analytics is at the core of clinical and translational informatics [1]. Data analysis involves several preprocessing activities such as data exploration, collection, cleaning, curation, interpretation, etc. These activities are a necessary prerequisite to perform analyses and answer research questions, and these activities are time-consuming in current research evnviornments. They can be broadly classified into the steps of data discovery, interpretation, and integration. We identify the need for a data fusion platform that will aid researchers in this time-consuming process.

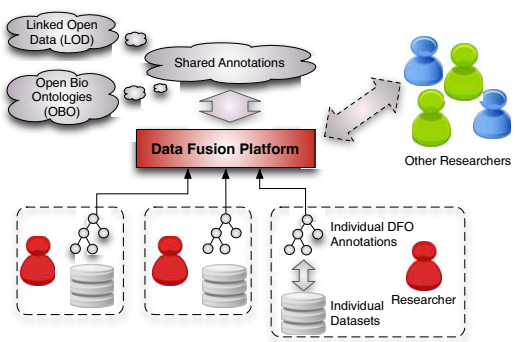


Figure 1– Role of Data Fusion Platforms in research

We define a **data fusion platform** as a system that provides automated, end-to-end, real-time and ad-hoc functionality across the data preprocessing tasks of data discovery, interpretation and integration. As shown in Figure 1, such a platform will serve as a middle-layer between researchers and crowd-sourced data collections promoting faster discovery and interpretation of datasets. It would allow researchers to publish their datasets by leveraging reusable annotations to standard reference ontologies and vocabularies.

Methods

To determine the plausibility of this notion, we conducted an informal interview with researchers, statisticians, and business analysts. To further understand the exact requirements of the researchers and assess the capabilities of the current set of tools to satisfy them, we conducted semi-structured interviews containing 8 questions with 4 biomedical researchers with varying levels of experience working, primarily, in different areas of research. In addition we also interviewed 2 business data analysts and 2 biostatisticians.

Results

The interviews revealed major differences in the process and philosophy of the individual researchers and the large-scale business analytics. About two-thirds of data driven research in biomedical informatics relies on tabular data. Currently, researchers have limited access to automated tools to help in the data preprocessing and usually perform all tasks manually.

Discussion

Informal interviews of researchers, statisticians, and business analysts confirmed that interactive data fusion platforms could serve to bolster the efficiency of multi-dimensional research. Even though existing tools and methods perform individual data preprocessing tasks well, none of the approaches are able to provide an end-to-end, interactive data fusion platform capable of performing in real-time. We argue that a data fusion platform with such capabilities is crucial for the effective research in the modern data environment.

The next step is to formalize and conduct interviews with a larger, more diverse set. The primary focus of ongoing research is to formally articulate the specifics of such systems and the frameworks that would be needed to design and implement them.

References

[1] Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc 16. 2009;3:316-27.

Address for correspondence
Satyajeet Raje, MS, PhD candidate
Depatment of Biomedical Informatics
The Ohio State University, Columbus, OH, USA
raje.3@osu.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1108

Towards a Formal Representation of Processes and Objects Regarding the Delivery of Telehealth Services: The Telehealth Ontology (TEON)

Filipe Santana[a], Stefan Schulz[b], Amadeu Campos[c], Magdala A. Novaes[c,d]

[a] *Centro de Informática (CIn), Universidade Federal de Pernambuco (UFPE), Recife, Brazil*
[b] *Institut für Medizinische Informatik, Statistik und Dokumentation (IMI), Medizinische Universität Graz, Austria*
[c] *Núcleo de Telessaúde (NUTES), Hospital das Clínicas, Universidade Federal de Pernambuco (UFPE), Recife, Brazil*
[d] *Departamento de Medicina Clínica, Centro de Ciências da Saúda (CCS), Universidade Federal de Pernambuco (UFPE), Recife, Brazil*

Abstract

This study introduces ontological aspects concerning the Telehealth Ontology (TEON), an ontology that represents formal-ontological content concerning the delivery of telehealth services. TEON formally represents the main services, actors and other entity types relevant to telehealth service delivery. TEON uses the upper level ontology BioTopLite2 and reuses content from the Ontology for Biomedical Investigations (OBI). The services embedded in telehealth services are considered as essential as the common services provided by the health-related practices. We envision TEON as a service to support the development of telehealth systems. TEON might also enable the integration of heterogeneous telehealth systems, and provide a base to automatize the processing of telehealth-related content.

Keywords:

Telehealth; Biological Ontology; Health Services.

Introduction

The lack of communication between telehealth systems creates (mostly) repetitive data, without any standardisation of the key concept meanings. We therefore recommend the use of ontologies and/or terminologies as a way to enrich the descriptions and content of telehealth-related systems and applications with domain-specific, controlled terms. According to this requirement, the aim of this study is to introduce ontological aspects concerning the Telehealth Ontology (TEON).

Methods

TEON was created using Description Logics (DL) [1], and implemented using the Web Ontology Language v2 (OWL2), built and edited via Protégé v5. TEON expands the upper domain ontology BioTopLite2 (BTL2) [2] and reuses content from the Ontology of Biomedical Investigations (OBI) [3], e.g., the class service that has TelehealthService as a subclass. TEON is available at http://www.nutes.ufpe.br/teon.

Results

TEON formulates classes and axioms to represent the delivery of telehealth services. It includes the description of actors, synchronicity profiles, health teams, and specific processes. Actor is a term related to the bearer of roles during the delivery of services. From telehealth services, there are three main roles: **requestor** (*RequestorRole*); **teleconsultant** (*TeleconsultantRole*); and **manager** (*ManagerRole*).

To express the synchronous profile of services, we took the notion of *PointInTime* and *TimeInterval* from BTL2. When processes are synchronous, their time interval and interval boundaries coincide. Otherwise they are asynchronous.

The notion of service, taken from OBI, describes them as "*planned processes in which two different entities are bearers of consumer and provider roles*" [3]. Telehealth services are delivered for different healthcare specialties such as telecardiology, teledermatology and telepsychiatry. These services are constrained by how they are delivered, e.g., via teleconsultation, second opinion, telediagnosis, among others.

Conclusion

In the current study, we described the formal description of the telehealth services domain in the Telehealth Ontology (TEON). Services that instantiate TelehealthService can be considered to have similar characteristics as common health care services. In addition, the roles played by the actors, the temporal dimensions (synchronicity and asynchronicity), are represented. As an ontology, we envision TEON's use as the formal background to guide the development of telehealth applications and systems. Ontology use in the telehealth domain might overcome the lack of consensus surrounding the terminology and subtleties about what is a service and what is telehealth practice.

References

[1] Baader F, McGuinness DL, Nardi D, Patel-Schneider P, Calvanese D, McGuinness DL, et al. The Description Logics Handbook: Theory, implementation, and applications [Internet]. 2nd ed. Cambridge University Press, 2003.
[2] Schulz S, Boeker M. BioTopLite: An Upper Level Ontol ogy for the Life Sciences. Evolution, Design and Application. Informatik 2013. Koblenz: IOS Press, 2013.
[3] Courtot M, Bug W, Gibson F, Lister AL, Schober D, Brinkman RR, et al. The OWL of Biomedical Investigations. Proceedings of the 5th OWLED. Karlsruhe, Germany; 2008.

Address for correspondence

Magdala A. Novaes – magdala.novaes@nutes.ufpe.br

Av. Prof. Moraes Rego, s/n, Cidade Universitária, Recife-PE, CEP 50.670-420, Brazil

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1109

Characterizing Health Information for Different Target Audiences

Yueping Sun[a], Zhen Hou[a], Li Hou[a], Jiao Li[a]

[a] *Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China*

Abstract

Different groups of audiences in health care: health professionals and health consumers, each have different information needs. Health monographs targeting different audiences are created by leveraging readers' background knowledge. The NCI's Physician Data Query (PDQ®) Cancer Information Summaries provide parallel cancer information and education resources with different target audiences. In this paper, we used targeted audience-specific cancer information PDQs to measure characteristic differences on the element level between audiences. In addition, we compared vocabulary coverage. Results show a significant difference between the professional and patient version of cancer monographs in both content organization and vocabulary. This study provides a new view to assess targeted audience-specific health information, and helps editors to improve the quality and readability of health information.

Keywords:

Target Audience; Element Level; Health Information; Readability.

Introduction

The National Cancer Institute's (NCI) Physician Data Query (PDQ®) Cancer Information Summaries are widely recognized as an important cancer information and education resource, with multiple target audiences [1]. Although general guidelines for the PDQ editors have been provided, no other measurable quality evaluations are available. Existing studies show that it is critical to precisely define measurable data attributes to value the quality of health data. In this paper, we aim to quantitatively compare the contents targeting professionals and patients at both element and vocabulary levels.

Materials and Methods

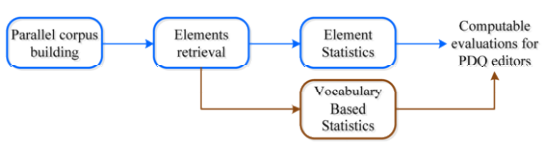


Figure 1– Steps of the characterizing method

The study method consists of 4 steps: 1) building a parallel PDQ corpus that includes both summaries targeting the health professionals and summaries targeting patients; 2) retrieval of elements; 3) statistical comparison of the elements, which returns computable evaluations for PDQ editors; 4) open-access and collaborative (OAC) consumer health vocabulary (CHV) based statistics.

Results

For the element level statistical analysis, we grouped the elements to 4 classes: 1) elements unique in the professional version (Unique in Pro); 2) elements which are remarkably more in the professional version (More in Pro); 3) elements which have similar distributions (Similar in Pro and Pat); 4) elements which are remarkably more in the patient version.

Table 1– Element Classification Results According to Statistics

Classes	Elements
Unique in Pro	LOERef; Refidx; Citation
More in Pro	Table; SummaryRef; Title; Para
Similar in Pro and Pat	ListItem; ItemizedList; SummarySection; ExternalRef; TermRef.
More in Pat	KeyPoint; MediaLink; GrossTermRef

Our results are consistent with the regulations given to PDQ editors. No citations have been found in the patient versions while more illustrations, represented by elements in "More in Pat" have been found. Besides, it is shown that the patient version has no level of evidence information (LOERef). Also we have obtained element distributions for both versions to statistically measure new batches.

The frequencies of an OAC CHV term in the two versions were compared. Considering that the two versions have different sizes, we assign the frequency of a term as 1 when the term appears more than one time in the summary. The overall frequency in the patient version is 260292, while the one in the professional version is 423739. It is shown that the patient targeted summaries cover far fewer OAC CHV terms than the professional ones.

Discussion and Conclusion

We introduced a new perspective of evaluating health information targeting professional and patients from the view of content organization and vocabulary usage. Our study provides quantitative guideline for monograph editors or health information authors to improve composition readability. In future work, context comparisons via Natural Language Processing (NLP) techniques will be explored.

References

[1] Richard E. Manrow, Margaret Beckwith, Lenora E. Johnson. NCI's Physician Data Query (PDQ®) Cancer Information Summaries: History, Editorial Processes, Influence, and Reach. J Canc Educ, 2014; 29:198–205.

Address for correspondence

Jiao Li, Ph.D., Email: li.jiao@imicams.ac.cn

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1110

A Pilot Ontology for Healthcare Quality Indicators

Pam White[a] and Abdul Roudsari[b]

[a] *School of Informatics, City University London, United Kingdom*
[b] *Department of Health Information Science, University of Victoria, British Columbia, Canada*

Abstract

Computerisation of quality indicators for the English National Health Service currently relies primarily on queries and clinical coding, with little use of ontologies. We created a searchable ontology for a diverse set of healthcare quality indicators. We investigated attributes and relationships in a set of 222 quality indicators, categorised by clinical pathway, inclusion and exclusion criteria and US Institute of Medicine purpose. Our pilot ontology could reduce duplication of effort in healthcare quality monitoring.

Keywords:

Ontologies; Quality Indicators, Healthcare

Introduction

Quality indicators are useful tools for monitoring healthcare outcomes. When quality indicators are issued by more than one governing body, they may overlap in content. Ontologies can facilitate automated quality monitoring by categorising and establishing relationships between concepts. In England's National Health Service (NHS), quality indicators are frequently measured electronically by using separate queries and data extraction for each quality indicator. An ontology for quality indicators issued from different sources can reduce effort needed to find data for quality indicators by linking common criteria in the indicators.

Method

We developed a pilot ontology that specifies inclusion and exclusion criteria, along with relationships between quality indicators. We categorised a large set of quality indicators [1] by clinical pathway, dimension, and US Institute of Medicine [2] purpose. The Clinical Pathway and Dimension categories were based on a 2008 strategic report for the NHS [3]. We used Statement and Definition metadata from the NHS [4] to specify layers of inclusion and exclusion criteria, creating chunks of indicator components. We then identified broader, narrower, and same level relationships between components of different indicators. We explain further detail about conceptualisation of the ontology in a separate article [5].

Results

Classes and Subclasses

Figure 1 shows the main classes and one of their subclasses.

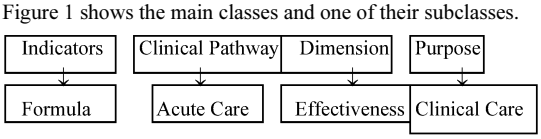


Figure 1. Examples of Classes and Subclasses

Properties

39 properties were assigned to the classes and/or subclasses. The Indicators class has 29 properties, with a maximum depth of seven layers of Inclusion/Exclusion Criteria (an exmaple of properties, due to different characteristics and names of the criteria).

Discussion

Our ontology was intended to make components of the indicators searchable, with a potential to reduce duplication of effort in finding data for common components of quality indicators from different sources. A use case example would be a clinical auditor seeking blood pressure measurement records within the past 9 months, for patients with hypertension. The auditor could search the ontology, using the keyword 'hypertension' to find related indicators. For example, they would find an indicator requesting records of lifestyle advice given to patients with hypertension. They would only need to search for records of patients with hypertension once to compile data for different indicators.

Conclusion

Our ontology offers the ability to search components of quality indicators from different sources, with a view to reducing duplication of effort in gathering data for indicators with common criteria. Although the ontology is useful in its present form, it would benefit from revision to improve its structure and to reduce the number of properties.

References

[1] NHS Information Centre. Indicators for quality improvement: Full indicator list. 2009. Available from: https://mqi.ic.nhs.uk/IndicatorsList.aspx. [Accessed 4 July 2012]

[2] Field M and Lohr K (Eds). Guidelines for clinical practice: from development to use. Institute of Medicine, Washington, D.C. National Academy Press. 1992.

[3] Darzi A. High quality care for all. NHS next stage review: Final report. London. Department of Health. 2008.

[4] NHS Information Centre Indicators for Quality Improvement. 2009. Available from: https://mqi.ic.nhs.uk/Default.aspx [Accessed 22 December 2014].

[5] White P and Roudsari A. Conceptual Analysis of a Diverse Set of Healthcare Quality Indicators. *Stud Health Technol Inform* 2015: 208:347-51.

Address for correspondence

Pam White, pamela.white.1@city.ac.uk

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1111

Fuzzy-Arden-Syntax-based, Vendor-agnostic, Scalable Clinical Decision Support and Monitoring Platform

Klaus-Peter Adlassnig[a,b,*], Karsten Fehre[b], Andrea Rappelsberger[a]

[a] *Section for Medical Expert and Knowledge-Based Systems, CeMSIIS, Medical University of Vienna, Vienna, Austria*
[b] *Medexter Healthcare GmbH, Vienna, Austria*

Abstract

This study's objective is to develop and use a scalable genuine technology platform for clinical decision support based on Arden Syntax, which was extended by fuzzy set theory and fuzzy logic. Arden Syntax is a widely recognized formal language for representing clinical and scientific knowledge in an executable format, and is maintained by Health Level Seven (HL7) International and approved by the American National Standards Institute (ANSI). Fuzzy set theory and logic permit the representation of knowledge and automated reasoning under linguistic and propositional uncertainty. These forms of uncertainty are a common feature of patients' medical data, the body of medical knowledge, and deductive clinical reasoning.

Keywords:
Arden Syntax;fuzzy methodologies; clinical knowledge representation;health IT integration and application.

Introduction

Arden Syntax is a medical knowledge representation and processing scheme for clinical decision support (CDS) systems. It originated in 1989 at a gathering of several medical informatics specialists from the USA, the Netherlands, and Sweden, at the Arden Homestead Retreat in Orange County, NY. The latest HL7- and ANSI-certified release is Arden Syntax version 2.10, which was approved in October 2014 [1]. This version consists of an augmented Arden Syntax, completely extended by fuzzy methodologies and ArdenML, an XML representation of the Arden Syntax code.

Methods

Medical logic modules (MLMs) are the basic representation and processing units in Arden Syntax. To execute MLMs written in Arden Syntax, one needs to write an interpreter or compiler for Arden Syntax, and an execution environment to process the MLMs. In addition, an authoring tool containing an editor for writing MLMs—which includes an execution engine for testing them before they become enacted—needs to be provided with such a suite of Arden Syntax software.

Following current software architectures and providing the Arden Syntax execution rule engine within a service-oriented architecture make it possible to offer CDS systems for a variety of tasks. In addition, clinical linguistic uncertainty can now be modeled using fuzzy sets; and propositional uncertainty can be expressed by rule-associated truth values. The computed results are then propagated and aggregated by fuzzy logic.

Results

We developed an authoring and testing environment, including an Arden Syntax version 2.10 compiler to write MLMs and compile and test the environment immediately. Based on an Arden Syntax rule engine, which executes the compiled MLMs, an Arden Syntax server is built around this engine to enable service-oriented access to and from client applications. To connect the server with host systems and data sources, three basic forms of technical integration were established:

1. *Web services for calling and data*: MLM and event calls are realized by SOAP or RESTful web services, with the service call also transferring the necessary data required for MLM processing.
2. *Web services for calling and server/database connector*: The second form of interconnecting is to call MLMs and events through SOAP or RESTful web services, but to access patient data directly from data sources through a so-called server/database connector (being an add-on to the Arden Syntax server).
3. *Data warehouse + rule engine = autonomous clinical decision support system*: Here the Arden Syntax server—including its rule engine and database connector—accesses a project-specific data warehouse (which can be quite general and extensive). This data warehouse receives "raw" patient data through communication servers or import routines from any external data source (in, e.g., HL7/XML/SQL format).

Conclusion

This technical solution was shown to be deployable in connection with hospital and intensive care information systems and with smart phone apps and research databases. The solution proved useful in a number of clinical fields.

Reference

[1] Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=372.

* Correspondence to klaus-peter.adlassnig@meduniwien.ac.at.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1112

Online Training Assessment For Primary Care Professionals Of The City Of São Paulo

Debora Cristina Alavarce[a], Heloisa Helena Ciqueto Peres[b]

[a] *School of Nursing, USP, São Paulo, Brazil*
[b] *Dep. ENO, School of Nursing, USP, São Paulo, Brazil*

Abstract

Students who participate in courses and training offered in distance mode receive positive results in learning evaluations. Most studies show improvement in knowledge in pre- and post- course comparisons; however, there are still few evaluations of long-term knowledge retention. This study compares the learning scores of online training for health professionals in three different times: before training, immediately after, and 180 days later. This technological production research includes quantitative, descriptive and multivariate correlational, longitudinal, and panel types. The results show that online training has a positive impact on learning and knowledge retention. Whereas learning can be related to the impact of training, i.e., with the transfer of what was gathered for professional practice, learning retention of online training applied to health professionals represents an important step and assists in identifying the best strategies so expected results are achieved.

Keywords:

Service training;Distance education; evaluation;Evaluation of learning.

Introduction

Online training includes a set of technologies and media, organized based on theoretical assumptions of education and communication. The present study adopted the theory of meaningful learning in order to support an analysis of the learning process and knowledge retention of online training for health professionals.

The learning evaluation is an important tool for estimating the adequacy of procedures and educational strategies used in training and to make it possible to assess whether the individual in fact learned what was taught. The evaluation serves also as a tool to demonstrate the effectiveness of the training and subsidizes the perfection of training itself, as well as needs assessment, evaluation of results, assisting managers and organisers of training to control the quality of their actions [1].

Materials and Methods

This study deals with the technological production linked to research foundations and management practice in nursing and health of the graduate program in nursing management (PPGEn) of the nursing school at the University of São Paulo (EEUSP). This group includes studies and research of information technology in work processes in nursing (GEPETE/EEUSP).

The technological product developed in this study was an online training on the topic identification of dementia in the elderly. The training was organized into 4 units with a schedule provided of 20 hours, the LMS used was Moodle ®. The units were developed using different learning digital objects. Each module contained a different activity, such as discussion forums, case study and practical cognitive assessment activity.

The study site was a specialized outpatient health services for the elderly located in a region of São Paulo / BR and students were graduates health profissinais who worked in the basic health units of the region.

Results

The learning evaluation applied in three different time frames was characterized as shown in Table 1:

- The assessment of learning online pre-training presented an average mark of 6.24 + 1.25. (Pre)
- The immediate evaluation was significantly higher with an average of 7.66 + 1.21 with an estimated gain of 1.41 points on the first note. (Pós1)
- Late Review, which kept an average of 7.24 + 0.84, and estimated gain of this compared to pre-training assessment online was 0.96 points. (Pós2)

Table 1 - Distribution of Average and Standard Deviation and Notes Obtained by Pupils

Type	N	Min	Max	Average	SD
Pre	40	3,66	8,68	6,24	1,25
Post 1	39	5,33	10	7,66	1,21
Pos 2	37	5,84	10	7,24	0,84

The survey results demonstrate that there has been a significant gain in immediate and long-term review when compared with the pre training notes. The Group proved to be heterogeneous in both the assessment moments that can be perceived by maintaining the standard deviation in 1.2.

References

[1] Queiroga F, Andrade JMd, Walnut R, Abbad GDS. learning Measures in TD&E theoretical foundations and methodological. In: Abbad GS, Mourão L, Markovic PP, Zerbini T, Borges-Andrade JE, veerababu R, editors. Evaluation measures in training, development and education: tools for managing people. 1 Ed. Porto Alegre: New Haven; 2012. p. 106-24.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1113

Allergy Risk Finder: Hypothesis Generation System for Allergy Risks via Web Service

Eiji ARAMAKI, Ph.D.[a], Shuko SHIKATA[a], Eriko WATABE[b], Mai MIYABE, Ph.D.[a], Yasuyuki USUDA[c], Satsuki AYAYA[d], Shinichiro KUMAGAYA, Ph.D. M.D.[d]

[a] *Design Unit, Kyoto University* [b] *IR-Advanced Linguistic Technologies Inc.*
[c] *The Graduate School of Human Environment of Sciences, Kyoto University*
[d] *Research Center for Advanced Science and Technology, The University of Tokyo*

Abstract

This study's aim was to build a web service that automatically collects and tests hypotheses for possible allergy risks. We crowdsourced for unknown allergy risks, and obtained odds ratios. By using the collected hypotheses, we built a web service that estimates allergy risks from a questionnaire (consisting of 10 questions that we gathered from the crowdsourcing task), and at the end, we asked the users their new hypotheses on possible allergy risks. The web service also asked the users to send their original hypotheses to contribute to find the cause of allergy. In the near future, clinical trials to validate the hypotheses found in this study are desired.

Keywords:

Crowdsourcing;Web service;Hypothesis generation;Allergy risk; e-health.

Introduction

This study generates clinical hypotheses by asking the public through our original website, the *Allergy Risk Finder*. Hypothesis generation is one of the most essential tasks for clinical research, and several studies have attempted to crowdsource such tasks [1-4]. This study used possible risks brought up by crowdsourcing tasks, and incorporated the risks in the seed questions All processes in this research were via the *Allergy Risk Finder*.

Materials and Methods

We set the known risks using review paper survey and the governmental guideline for allergy [5]. We extracted 29 hypotheses, and 8 of them were used for the system as the seed questions. The other 21 were used for evaluation. Firstly, the users filled in the questionnaire items (hereafter, questions), and the questions changed as the number of users increased. The questionnaire consisted of three types of questions: (1) Profile questions obtained the profile of the user; having allergy or not, his/her gender and age, (2) Risk questions (contained seed and newly proposed questions by users in (3)); asked the user's environmental situation by randomly selecting the allergy hypotheses from the risk pools, and (3) Original risk-proposing questions survey to propose novel allergy risks (1 to 5 answers required); used as the questions for the next and further round.

We repeated this procedure five times. Potentially promising top 100 hypotheses with high odds ratios were kept in the allergy risk pool, and the other hypotheses with lower potential were discarded.

Results

Since this service has launched, the *Allergy Risk Finder* has already found more than 152 new hypotheses of allergy risks. We evaluated the performance of the already-known risks with the following 2 aspects: (1) the propotions of already-known risks found within the users' answers, and (2) we used 21 already-known risks to evaluate the propotions of hypotheses produced by users with high odds ratios (over 1.0 in lower bound of 95% confident interval). The 152 newly proposed risks included all 21 already-known risks, which suggested the broad coverage of the *Allergy Risk Finder*. Among all newly proposed hypotheses of allergy risks, 43 hypotheses showed odds ratios of over 1.0. It was notable that these included 13 already-known risks, indicating the validness of this process.

Conclusion

The *Allergy Risk Finder* could show the whole allergy research process from hypothesis generation to questionnaire-based validation. Clinical trials to validate the hypotheses found in this study are desired.

References

[1] Swan M (2012) Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 14: e46.

[2] Bongard J, Hines P, Conger D, Hurd P, Lu Z (2012) Crowdsourcing Predictors of Behavioral Outcomes. Systems, Man, and Cybernetics: Systems, IEEE Transactions 43: 1-10.

[3] Bevelander KE, Kaipainen K, Swain R, Dohle S, Bongard JC, et al. (2014) Crowdsourcing novel childhood predictors of adult obesity. PLoS One 9: e87756.

[4] Armstrong AW, Harskamp CT, Cheeney S, Schupp CW (2012) Crowdsourcing in eczema research: a novel method of data collection. J Drugs Dermatol 11: 1153-1155.

[5] Jenerowicz D, Silny W, Danczak-Pazdrowska A, Polanska A, Osmola-Mankowska A, et al. (2012) Environmental factors and allergic diseases. Ann Agric Environ Med 19: 475-481.

Address for correspondence

Eiji ARAMAKI eiji.aramaki@gmail.com
Center for the Promotion of Interdisciplinary Education and Research, Kyoto University. Kyoto 600-8815, Japan

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1114

Toward a Global eHealth Observatory for Nursing

Claudia C. Bartz[a], Nicholas R. Hardiker[b], Amy Coenen[c]

[a] *International Council of Nurses, Geneva, Switzerland*
[b] *School of Nursing, Midwifery, Social Work & Social Sciences, University of Salford, UK*
[c] *College of Nursing, University of Wisconsin-Milwaukee, Milwaukee, WI, USA*

Abstract

This poster summarizes a review of existing health observatories and proposes a new entity for nursing. A nursing eHealth observatory would be an authoritative and respected source of eHealth information that would support nursing decision-making and policy development and add to the body of knowledge about professional nursing and client care outcomes.

Keywords:

Observatory; eHealth; Nursing.

Introduction

eHealth is the use of information and communication technologies (ICT) for health. The aim of knowing who, what, when, where, why and how of "eHealth" has resulted in the emergence of eHealth observatories. Professional observatories are essentially institutions or entities that monitor and evaluate a selected topic over time. The health community is trying to more deliberately describe and advance its eHealth universe.

A global eHealth observatory for nursing is needed. The purpose of this poster is to propose an international eHealth observatory for nursing that would consider nursing eHealth roles, projects and research that would contribute to knowledge about professional nursing and client care outcomes in eHealth environments.

Methods

The work was carried out under the auspices of the International Council of Nurses (ICN) eHealth Programme.

First, using descriptive methods, published and web-sourced materials of health-related observatories were examined for characteristics applicable to a nursing eHealth observatory.

Second, a multidisciplinary team of eHealth and policy experts was tasked by ICN to envision an eHealth observatory for nursing and provide input into the planning process. The team was given guiding principles for their work, including the ICN Strategic Plan and the eHealth Programme goals and outcomes.

Results

Four health-related observatories were examined, with attention to their purposes and scopes. These were the (a) WHO Global Observatory for eHealth, (b) University of Victoria eHealth Observatory, (c) Women Observatory for eHealth (WeObservatory) and (d) ICN Observatory on Licensure and Registration. The observatories varied but they were clear in their aim of serving as a central hub for their selected target topic in health: global data, information and knowledge; information systems; women's health and well-being; and nursing regulation, respectively. Two more health-related observatories could be useful to an international eHealth observatory for nursing: (a) the International Care Ethics Observatory and (b) the Eastern Region Public Health Observatory.

The team was very enthusiastic about an eHealth observatory for nursing as a means to identify, disseminate and respond to trends in eHealth. An observatory would influence policy and decision-making and serve as an incubator for projects going forward. It could bring a unique nursing perspective and voice to the international eHealth community. Its key characteristics would be neutrality and objectivity. The observatory would aim to be an authoritative and respected source of recommendations and knowledge in matters related to nursing and eHealth. The observatory could identify best practices for health promotion, disease prevention and care delivery in ICT-enabled environments.

The team recommended that stakeholders include the National Nurses Associations (NNA) who are members of ICN, nurses worldwide, non-governmental organizations which are involved in eHealth or other health-related endeavors, multidisciplinary professional health organizations, educators and researchers. Other stakeholders could include citizens and communities with health care needs, legislators, health economists and commercial entities.

Conclusion

The proposed observatory would bring a strong, focused nursing perspective to the eHealth environment. Outcomes could include improved policy, practice and management of eHealth in nursing through availability of relevant, timely and high-quality evidence and information; new knowledge that would contribute significantly to nursing and health through the use of ICT; and empowerment and engagement of nurses worldwide through data-based information and evidence for practice. A strategic plan and operational processes are the next steps toward a nursing eHealth observatory.

Address for correspondence

Claudia C Bartz, cbartz@uwm.edu

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1115

Representation of Biomedical Expertise in Ontologies: a Case Study about Knowledge Acquisition on HTLV viruses and their clinical manifestations

Kátia Cardoso Coelho[ac], Maurício Barcellos Almeida[b]

[a] *PhD Candidate, Federal University of Minas Gerais, Belo Horizonte, Brazil*
[b] *Associate Professor, Federal University of Minas Gerais, Brazil*
[c] *Hemominas Foundation, Belo Horizonte, Brazil*

Abstract

In this paper, we introduce a set of methodological steps for knowledge acquisition applied to the organization of biomedical information through ontologies. Those steps are tested in a real case involving Human T Cell Lymphotropic Virus (HTLV), which causes myriad infectious diseases. We hope to contribute to providing suitable knowledge representation of scientific domains.

Keywords:
Knowledge acquisition, Ontologies, Knowledge representation.

Introduction

This study investigates the activity of Knowledge Acquisition (KA) within the scope of biomedicine. In order to improve that activity, we propose procedures for knowledge acquisition, which adhere to some of the best practices found in the literature [1][2][3][4]. We systematize these procedures in a list of methodological steps with the aim of testing their feasibility in a real case.

Materials and Methods

The methodological steps are developed from a comprehensive literature review and tested in a real case of knowledge acquisition about HTLV.

Methodological steps for KA

The first phase is the *survey phase,* the second was called *elicitation phase*, consisting of interviews and applying KA techniques to experts. The three major stages that comprise that cycle are: *etiological process, course of disease* and *therapeutic response.* The first stage called *etiological process*, The *course of disease* phase starts with the clinical manifestation of a disease. At this moment, the disorder manifests itself through symptoms, which the patient is able to identify.

The clinical framework is composed of symptom representation records, as well as physical and laboratory exam results.

Results

After the preliminary organization of the terms, the results were presented to the main researcher, in order that she could validate them. In this step, the expert had to accept or not accept what was presented or suggest changes, which would be recorded for new future evaluations. All terms and definitions were accepted by the specialist.

Conclusion

The KA activity from experts as part of the process for developing ontologies can also be understood as a preliminary activity before automatic term extraction. Specialists are required to judge whether the extracted terms make sense in the domain. The biomedical vocabulary we come up with also has a relevant function: consensually define the meaning of terms used in medical practice and research. This is made possible by directly considering knowledge acquisition from experts.

Acknowledgments

We thank the *Hemominas* Foundation, *Minas Gerais*, Brazil and the Interdisciplinary HTLV Research Group (GIPH).

References

[1] Hayes-Roth F, Waterman DA, Lenat DB. Building Expert Systems. Massachusetts: Ed. Addison-Wesley. 1983, 350 p
[2] Milton NR. Knowledge acquisition in practice: a step-by-step guide. Cranfield: Springer, 2007. 176p.
[3] Milton N, Clarke D, Shadbolt N. Knowledge engineering and psycology: Towards a closer relationship**.** *Int J Hum Comput Stud* 2006; 64:1214-1229. Available from: http://www.sciencedirect.com/science/article/pii/S1071581906001212
[4] Boose JH, Gaines BR. Knowledge Acquisition for Knowledge-Based Systems: Notes on the State-of-the-Art. *Mach Learn* 1989; 4:377-394. Available from: http://www.springerlink.com

Address for correspondence

Kátia C.Coelho email katiaccoelho@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1116

Trigger Development for the Improvement of Neurological Patient Care

Eija Kivekäs[a], Ulla-Mari Kinnunen[b], Kaisa Haatainen[c,e], Reetta Kälviäinen[d,e], Kaija Saranto[b]

[a] *Research Centre for Comparative Effectiveness and Patient Safety, University of Eastern Finland, Kuopio, Finland*
[b] *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*
[c] *Department of Nursing Science, University of Eastern Finland, Kuopio, Finland*
[d] *Department of Neurology, University of Eastern Finland, Kuopio, Finland*
[e] *Kuopio University Hospital, Kuopio, Finland*

Abstract

By analyzing medical records, we developed triggers for epilepsy patients' care coordination. Thirteen triggers with potential to affect patient care outcomes and safety were found.

Keywords:

Patient Safety; Documentation; Epilepsy.

Introduction

A trigger can be defined as a patient-specific indicator highlighting changes in patient's well-being or a clue to identify adverse events. The Global Trigger Tool (GTT) method has been used commonly in patient safety research [1]. Patient safety refers to patients who are protected from avoidable harm.

In Europe, epilepsy affects 6 million people of all ages. The lifetime risk of epilepsy is 2%. The etiology of epilepsy is a major determinant of a patient's clinical course and prognosis [2]. Documentation of patient care is fundamental and critical to communicate a patient's individual needs and his/her responses to care. Electronic documentation will improve the quality of care provided to patients by standardizing both structures and content of Electronic Health Records. The value of structured language has been considered to support patient care workflow and the delivery of care. [3].

The purpose of our study is to recognize triggers in epilepsy patients by analyzing their medical records. The focus of our study is on the continuity of care and patient safety.

Methods

The pilot research data included the medical records of adult epilepsy patients (n=20) from a Finnish tertiary hospital (2004-2012). The study was a retrospective medical records' review of randomly selected adult epilepsy patients' records. Two primary reviewers performed a retrospective chart review and their work was validated by two secondary reviewers of a multidisciplinary team. The triggers were initially extracted from the medical records and after that these were checked from the literature.

Results

The medical records of patients suffering from unmanageable epilepsy consisted of 15 female and five male patients. Their age varied between 17 and 69 years. The length of patient care varied from six months to eight years, and the average length was five years. Thirteen triggers (Table 1) were identified for epilepsy patient care. The triggers provided meaningful information for patient care, care processes and patient safety of epilepsy patients. The preliminary findings of this study showed that the triggers could indicate changes in the epilepsy patients' health and well-being.

Table 1 – Epilepsy patient's triggers (n=13)

Triggers
Bowel symptoms
Cognition and motoric ability (motoric skills)
Eyesight problems
Headaches
Hormonal imbalance
Mental stress
Mood fluctuation
Quality of social life (coping skills)
Skin problems
Insomnia
Socioeconomic situation
Fatigue
Weight fluctuation

Discussion

Thirteen triggers were identified, reprensenting the diverse range of epilepsy patients' health and well-being. The triggers offer health care personnel an opportunity to develop epilepsy patients' care documentation. The development of documentation improves patients' care, the continuity of care between care providers, and patient safety.

Conclusion

This pilot study demonstrated the need for conducting a more comprehensive study with more substantial data.

References

[1] Institute for Healthcare Improvement (IHI). Introduction to Trigger Tools for Identifying Adverse Events. Cambridge, Massachusetts, USA, 2014.

[2] Perucca E, Glynn M, de Boer H, et al. Epilepsy Research Priorities in Europe. Available: http://www.ilae.org/Visitors/policy/documents/ERF-Report.pdf

[3] Kelley TF., Brandon DH., Docherty SL. Electronic Nursing Documentation as a Strategy to Improve Quality of Patient Care. Journal of Nursing Scholarship 2011; 43:2, 154-62

Address for correspondence

Eija Kivekäs, eija.kivekas@uef.fi

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1117

Patient Empowerment through Personal Medical Recommendations

Haridimos Kondylakis, Lefteris Koumakis, Eleni Kazantzaki, Maria Chatzimina, Maria Psaraki, Kostas Marias, Manolis Tsiknakis[a]

[a] *Computational BioMedicine Laboratory, FORTH-ICS, Heraklion, Crete, Greece*

Abstract

Patients today have ample opportunities to inform themselves about their disease and possible treatments using the Internet. While this type of patient empowerment is widely regarded as having a positive influence on the treatment, there exists the problem that the quality of information that can be found on online is very diverse. This paper presents a platform which empowers patients by allowing searching in a high quality document repository. In addition, it automatically provides intelligent and personalized recommendations according to the individual preferences and medical conditions.

Introduction & Methods

During the last decade, the number of users who look for health and medical information online has dramatically increased. However, despite the increase in those numbers and the vast amount of information currently available online, it is very hard for a patient to accurately judge the relevance of information to his own case. Although there are already several approaches trying to provide patients with search engines containing high quality medical information such as WebMD, MayoClinic Patient Care, Medicine Plus, HONSearch etc., these engines provide a rather limited set of information, and they are not dynamically adapted according to patient's preferences or medical history.

This paper focuses on current research activities related to the implementation of a Personal Medical Information Recommender (PMIR). PMIR (an early version has already been presented [1]) is targeted at improving the opportunities that patients have to inform themselves using the Internet, about their disease and possible treatments, and providing them with personalized information and recommendations.

Results & Conclusions

The PMIR is integrated into a PHR as a set of individual apps and services shown in Figure 1. Using the *Semantic Annotator app,* an expert is able to register external, high-quality web documents that contain useful information. Those documents are then annotated using the *Semantic Annotator Service* with terms from the SNOMED-CT, the LOINC and the RXTerms ontologies. All terms and documents are stored in the *Terms Indexing DB.*

The patient is able to select the *PMIR search app* to look for useful information. Besides searching for relevant results, results can be rated according to the patient's opinion. The clicks and the ratings of each user are stored in the *Patient Preferences* database. In addition, interesting web documents are automatically recommended to patients using the *Automatic Recommendation App*.

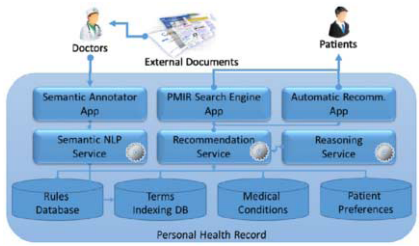


Figure 1 - PMIR Architecture

Both search results and automatic recommendations are provided through a *recommendation service* which uses a varation of the the vector space model. The service considers the following databases to make the results of the query as personalized as possible: (a) The *Patient Preferences Database* containing user preferences that are acquired as the patient browses the results presented to him; (b) The *Medical Conditions Database* including the medical conditions of the patient as they have been logged by the patient himself and annotated with the aforementioned ontologies; (c) The *Terms Indexing DB* with the annotations of the indexed documents; and (d) The *Rules Database* including rules that are used for generalizing patient queries. Rules are generated by experts (e.g. include specific subsumption relations) and are exploited in order to increase the recall of the user search.

To the best of our knowledge, PMIR is the only system exploiting patient's profiles to provide both automatic and non-automatic high quality information to patients employing semantics, reasoning and also exploiting user preferences.

References

[1] Kondylakis H, Koumakis L, Rüping S, Kazantzaki E, Marias K, Tsiknakis M. PMIR: A Personal Medical Information Recommender, MIE, 2014. pp. 1193.

Address for correspondence

Haridimos Kondylakis, N. Plastrira 100, Vasilika Vouton, Heraklion, Crete, Greece, GR-700 13, kondylak@ics.forth.gr

This work has been supported by the EU projects iManageCancer, EURECA and MyHealthAvatar and has been partially funded by the European Commission under contracts H2020-643529, FP7-288048 and FP7-600929.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1118

A software tool to analyze clinical workflows from direct observations

Marco Schweitzer[a], Nelia Lasierra[b], Alexander Hoerbst[a]

[a] *Research Division for eHealth and Telemedicine, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria*
[b] *STI - Semantic Technology Institute, University of Innsbruck, Innsbruck, Austria*

Abstract

Observational data of clinical processes need to be managed in a convenient way, so that process information is reliable, valid and viable for further analysis. However, existing tools for allocating observations fail in systematic data collection of specific workflow recordings. We present a software tool which was developed to facilitate the analysis of clinical process observations. The tool was successfully used in the project OntoHealth, to build, store and analyze observations of diabetes routine consultations.

Keywords:
Observation, process analysis, BPMN.

Introduction

Direct observation is a valuable method for collecting data about physicians' interactions with patients and others, such as the EHR, without depending on their direct collaboration. This method was selected to identify clinical workflows and processes related to IT-usage during the physician encounter as part of the OntoHealth project. Our goal was to identify clinical workflows to later enable their implementation through a computer system and make the interaction with the EHR easier for users. Observational workflow data needs to be gathered and stored in a convenient way to facilitate further analysis. However, existing tools (e.g. [1]) often fail at managing observational data in a versatile and flexible manner so that specific information about executed tasks and the interrelation among them may be handled. We developed a software tool that allows managing of context-related workflows, facilitating observation storage and analysis.

Materials and Methods

The literature review towards identifying workflows [2] led us to a tripartite categorization for generic actions, diabetes-specific data-elements and contexts that could be imported into the application for initial workflow categorization. In order to provide a graphical view and facilitate the analysis, we used the business process model and notation (BPMN) v2.0 for workflow visualization.

Results

The developed JAVA-application provides a SWING user-interface, enabling users to record and visualize observations regarding clinical workflows (Figure 1). The creation of a new observation as a business process diagram is realized using the open-source framework jGraphX (www.jgraph.com) tailored to the requirements of BPMN. A limited set of BPMN elements were integrated: users can create start/end events, tasks, gateways and connect all the elements to represent a sequence flow. Each task allows users to define properties of task-related characteristics (name, duration, IT-relation, executed person, etc.) that also comprise context-related information organized as a hierarchical categorization. The current version allows management of three categorizations. All emerging data is stored in a MySQL-database. BPMN- and context-related statistical values (e.g. process execution time, number of assigned elements) can be determined within the application and exported for further statistical analysis. Observations of diabetes routine examinations were documented using this application. During the input-process, categorizations were modified and final results were used for further analysis. This tool allowed for systematically identifying and quantifying the most common actions, data-elements and contexts used within diabetes routine examinations.

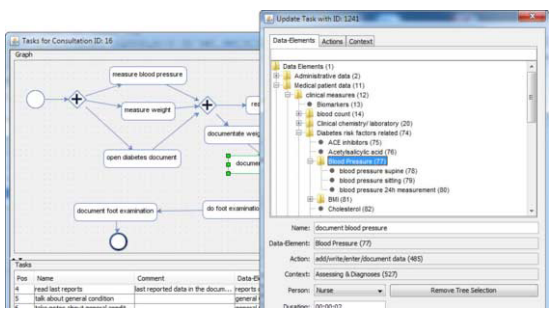


Figure 1: Screenshot of the application

Acknowledgments

This work has been supported by the Austrian Science Foundation (FWF), project number P 25895-B24.

References

[1] Westbrook JI, Ampt A. Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. Int J Med Inform 78 (2009): 25-33.

[2] Schweitzer M, Lasierra N, Oberbichler S, et al., Structuring clinical workflows for diabetes care - An overview of the OntoHealth approach, Appl Clin Inform 5 (2014), 512-526.

Address for correspondence

Marco Schweitzer: marco.schweitzer@umit.at

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1119

Knowledge-based immunosuppressive therapy for kidney transplant patients – from theoretical model to clinical integration

Walter Seeling[a], Max Plischke[b], Jeroen S. de Bruin[a], Christian Schuh[a]

[a] *Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*
[b] *Department of Internal Medicine III, Vienna General Hospital and Medical University of Vienna, Austria*

Abstract

Immunosuppressive therapy is a risky necessity after a patient received a kidney transplant. To reduce risks, a knowledge-based system was developed that determines the right dosage of the immunosuppresive agent Tacrolimus. A theoretical model, to classify medication blood levels as well as medication adaptions, was created using data from almost 500 patients, and over 13.000 examinations. This model was then translated into an Arden Syntax knowledge base, and integrated directly into the hospital information system of the Vienna General Hospital. In this paper we give an overview of the construction and integration of such a system.

Keywords:

Kidney Transplantation, Immunosuppression, Clinical Decision Support Systems.

Introduction

Patients suffering from (chronic) kidney diseases are often in need of a kidney transplant, after which they are required to follow a regiment of immunosuppresive therapies, which need to be tailored precisely to each patient; if not, the new organ could be rejected, or the patient becomes more susceptible to infections and other adverse medical events [1, 2].

To facilitate higher quality immunosuppresive therapy, a knowledge-based system (KBS) was developed for the manipulation of dosages for Tacrolimus, an immunosppressive agent often used after kidney transplants. A theoretical model was developed to classify the different types of kidney-transplant patient profiles [3]. The resulting profiles were then translated to a knowledge-base, and embedded directly in the hospital information system (HIS) where it is currently under clinical evaluation.

Methods

Data used to generate the theoretical model were historical patient data on immunosuppressive therapy and laboratory tests, gathered between 1995 and 2008. In total, 492 patients were included with 13,053 examinations. The theoretical model was built using conditional inference trees (CITs) [4], which created patient profiles, associated distributions and intervals of medication adaption (decrease, increase or maintain). The resulting classification model and distributions were then translated and implemented in Arden Syntax medical logic modules (MLMs) [5]. The graphical user interface (GUI), was created using Dynpro.

Results

The theoretical model resulted in 16 classes of patients and associated distributions, which were then translated to an MLM rule-base. A GUI was written to interact with the MLM, allowing physicians to input medication and immunosuppressant data directly from the HIS, and recalculate decisions at runtime. Based on the supplied parameters, the physician receives the patient distribution (percentage of patients with reduced, increased and unchanged medication), plus medication adaption intervals for each class. The results are accompanied by a written explanation for clarification.

Conclusion

Using electronically available patient and laboratory data, a method for semi-automated immunosuppressive therapy determination was created to guide nephrologists in the best course of action based on historical patient data. As this system is still under development, more rigorous validation must be done before the system can be widely distributed.

References

[1] S. Ho, et al., "The mechanism of action of cyclosporin A and FK506", *Clin Immunol Immunop,* vol 80(3), pp. S40-S5, 1996.

[2] S. Masuda, and K. Inui, "An up-date review on individualized dosage adjustment of calcineurin inhibitors in organ transplant patients", *Pharmacol Therapeut,* vol 112(1), pp. 184-98, 2006.

[3] W. Seeling, M. Plischke, and C. Schuh, "Knowledge-based tacrolimus therapy for kidney transplant patients", *Stud Health Technol Inform,* vol 180, pp. 310-4, 2012.

[4] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework", *J Comput Graph Stat,* vol 15(3), pp. 651-74, 2006.

[5] G. Hripcsak, "Writing Arden Syntax Medical Logic Modules", *Comput Biol Med,* vol 24(5), pp. 331-63, 1994.

Address for correspondence

Walter Seeling, M.Sc., Section for Medical Expert and Knowledge-Based Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Email: walter.seeling@medunwien.ac.at

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1120

Clinical application of the integrated multicenter discharge summary database

Suzuki Takahiro[a], Doi Shunsuke[b], Hatakeyama Yutaka[c], Honda Masayuki[d], Matsumura Yasushi[e], Shimada Gen[f], Takasaki Mitsuhiro[g], Tsumoto Shusaku[h], Yokoi Hideto[i], Takabayashi Katsuhiko[a]

[a]Department of Medical Informatics and Management, Chiba University Hospital,
[b]Department of Welfare and Medical Intelligence, Chiba University Hospital,
[c]Department of Medical Informatics, Kochi University Hospital,
[d]Department of medical informatics, Nagasaki University Hospital,
[e]Medical Informatics, Osaka University Graduate School of Medicine,
[f]St. Luke's International Hospital Medical Information System Center,
[g]Medical Information & Communication Center, Saga University Hospital,
[h]Department of Medical Informatics, Faculty of Medicine, Shimane University Hospital,
[i]Department of Medical Informatics, Kagawa University Hospital

Abstract

We performed the multi-year project to collect discharge summary from multiple hospitals and made the big text database to build a common document vector space, and developed various applications. We extracted 243,907 discharge summaries from seven hospitals. There was a difference in term structure and number of terms between the hospitals, however the differences by disease were similar. We built the vector space using TF-IDF method. We performed a cross-match analysis of DPC selection among seven hospitals. About 80% cases were correctly matched. The use of model data of other hospitals reduced selection rate to around 10%; however, integrated model data from all hospitals restored the selection rate.

Keywords:
Discharge summary, Text mining, Multi center Database.

Introduction

We have analyzed clinical documents and extracted necessary information using morphological analysis and a vector space model before. We presented the analysis of data collected from four hospitals at Medinfo2013. In this study, we collected data from seven hospitals and show more generalized results.

Materials and Methods

Data of present illness, past history, family history, and disease name from discharge summaries for the period from 2009 to 2011 at the following hospitals were output to an external file. The number of cases were 42,685 at Chiba University Hospital, 42,866 at Osaka University Hospital, 40,075 at Kagawa University Hospital, 10,629 at Kochi University Hospital, 41,465 at Nagasaki University Hospital, 28,704 at Saga University Hospital, and 37,483 at St. Luke's International Hospital, for a total of 243,907 cases. We used the term frequency-inverse document frequency (TF-IDF) method for text mining. Diagnosis Procedure Combination (DPC) selection was performed by extracting the DPC codes for at least 20 cases and randomly dividing them at a ratio of 7:3 into those used for model construction and those used for verification. The vector spaces of the model and the inputted summaries were compared, and the most similar vector was retrieved. We performed the cross-match analysis between 7 hospitals and integrated data of all hospitals. We built prototype of web application for similar-case retrieval system using this method.

RESULTS

In total, 27,220,461 index terms were extracted. The mean number of terms per summary varied widely, from 160 terms for Chiba University Hospital to 383 terms for Osaka University Hospital. Extremely short summaries were present at all of the hospitals. The distribution of summary lengths in Chiba University Hospital has two peaks, at 50 terms and 200 terms, whereas other hospital has normal distribution with a peak from 200 to around 300 terms. The number of index terms per summary was generally according to major diagnostic category (MDC). In all hospitals, a trend toward shorter summaries was seen for ophthalmology and otolaryngology diseases, whereas summaries of psychiatric and circulatory diseases were longer. The results of cross-matching test using 6 digits of the DPC code are shown in Table 1. For all of the hospitals, the use of own model data for each hospital indicated a high selection rate, with concordance of 77% to 85%. The use of model data of other hospitals reduced selection rate to around 10%, however, improvement of precision could be observed when we used integrated model data from all hospitals. Web application showed good response and accuracy (data not shown).

Table 1- Cross-matching results using 6 digits of DPC code

Model / Verification	Chiba Univ. Hospital	Kagawa Univ. Hospital	Kochi Univ. Hospital	Nagasaki Univ. Hospital	Osaka Univ. Hospital	Saga Univ. Hospital	St. Lukes Hospital	Integrated Data
Chiba Univ. Hospital	83%	75%	69%	77%	73%	77%	74%	83%
Kagawa Univ. Hospital	56%	78%	53%	66%	71%	65%	63%	74%
Kochi Univ. Hospital	75%	75%	80%	80%	75%	77%	73%	81%
Nagasaki Univ. Hospital	77%	78%	72%	85%	79%	82%	78%	85%
Osaka Univ. Hospital	72%	75%	69%	76%	84%	77%	76%	81%
Saga Univ. Hospital	72%	73%	66%	78%	74%	81%	78%	80%
St. Lukes Hospital	73%	72%	64%	77%	73%	80%	81%	77%

Discussion

The results of this study suggest the presence of common factors for diseases, and indicate a potential for a large text database with a common and general document vector space by integrating data from multiple hospitals. Use of this database not only provides improvements for the text mining accuracy, but also makes a contribution towards medical applications.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1121

Oncotherapy: A System for Requesting Chemotherapy Protocols

Laura Vera Righi[a]

[a] *Department of Medical Informatics, National Cancer Institute, Uruguay*

Abstract

A clinical decision support system is able to provide oncologists with suitable treatment options at the moment of decision making regarding which chemotherapy protocol is the best to apply to a particular oncological case. The National Cancer Institute has created a Guidelines Committee that establishes therapeutical options for each clinical case. The Health Informatics Department has developed Oncotherapy, a knowledge database that incorporates information provided by the Guidelines Committee. Oncotherapy includes a tailored information repository to provide oncologists in the public health system with the chemotherapy protocols available given three types of data: clinical diagnosis, clinical stage and therapy criteria. The protocol selected by the treating oncologist is sent back to Oncotherapy, which may create new knowledge that can be incorporated into the knowledge database. In this way, the system supports making the best decision according to the chemotherapy protocol options available. Furthermore, it can warn of errors that could result from mistakenly chosen therapies.

Keywords:

Clinical decision support system, Decision making, Chemotherapy protocols, Knowledge database.

Introduction

A clinical decision support system plays an important role in selecting a chemotherapy protocol. It helps physicians to select the best option for each particular case, drawing on knowledge of experts in the field and the opinions of the oncological community. Morover, this type of system can reduce the occurrence of errors in the physician's decision making process [1]. This type of error is not the most frequent, but surely one of the most serious, because in most cases it does not allow the possibility of correction before damage occurs. This oncology process also facilitates oncology patients' access to the same chemotherapy protocols based on their entry criteria. The objective of this study is to show a decision making process in the prescription of chemotherapy protocols that was assisted by a clinical decision support system called Oncotherapy.

Methods

The National Cancer Institute collects and processes by hand more than 2500 requests for chemoterapy protocols every month. Automation is an extremely challenging goal. We created an engine using guidelines for oncological therapies provided by the Guidelines Committee. This engine will provide the available chemotherapy protocols according to the following three types of data: clinical diagnosis, clinical stage and therapeutical criteria. The requests go through a designed workflow and may be automatically approved by the system if they match the criteria established by the engine. If a request does not match the criteria it is submitted to a Therapies Committee and may be approved or rejected. If the treating oncologist is unable to find the intended chemotherapy protocol option by using the engine, he or she can send a request to the Guidelines Committee, which may then enter new knowledge into the engine to enable the new option. In this way the system allows a dynamic interaction between the Guidelines Committee and the country's oncological community, producing feedback for the engine. A messenger service is provided in the system to help communication among oncological community members, as well as with the Guidelines Committee. This medical informatics solution will create a robust procedure for requesting oncology therapies according to guidelines in the National Health System. The interface is expressed in oncological language and uses color and images to help physicians reach the desired function promptly.

Results and Conclusion

Deciding on chemotherapy protocols for a patient is very challenging. In routine practice, physicians participate in Oncology Grand Rounds or Tumor Committees. This is not always possible in places far from oncology centers. This type of clinical decision support system should improve oncological assistance along the best decision making process, although the outcome of the application depends on the accurate and rigorous use of the informatics tools.

References

[1] Bates DW, Gawande AA. Patient safety: Improving safety with information technology. New Engl J Med. 2003 Jun.

Address for correspondence

lverarighi@gmail.com

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1122

Linked Health Data: how linked data can help provide better health decisions

Fernanda Farinelli[a], Maurício Barcellos de Almeida[b], Yóris Linhares de Souza[c]

[a, b] *School of Information Science, Federal University of Minas Gerais, Brazil.*
[c] *Institute of Technical Education of Minas Gerais, Brazil.*

Abstract

This paper provides a brief survey about the use of linked data in healthcare to foster better health decisions and increase health knowledge. We present real cases from the Brazilian experience and emphasize some issues in research. This paper is not intending to be fully comprehensive, we discuss some open issues and research challenges in linked data and the technologies involved. We conclude that even though linked data has been adopted in many countries, some challenges have to be overcome, for example, interoperability between different standards. A defined solution able to foster the semantic interoperability between different standards must be developed. Benefits contributed through linked health data involve better decision making on diagnostics, assertive treatments, knowledge acquisition, improvements in quality healthcare service to citizens.

Keywords:

Linked health data; linked data; ontology; decision making

Introduction

The World Health Organization (WHO) considers that health should be promoted by complete cooperation between individuals and states. The use of the Internet is an effective way of improving this cooperation. One alternative is Linked Data (LD) [2]. We present a brief discussion about the uses of LD and challenges in interoperability.

Materials and Methods

We conducted a bibliographic research using CiteSeerx, BioMed, ScienceDirect, Pubmed, ACM Digital Library, to mention a few. We selected cases that portrait the Brazilian reality.

Results

We surveyed some cases. First, a cohort study applied on data from different healthcare units through LD technology which contributed to assess prenatal care [3]. Secondly, semantically linked data about nursing found web aided in providing treatments [4].

Despite that LD research is increasing in the medical field, major challenges are still encountered. An error in data linkage can cause a decision making errors and consequently a faulty diagnosis which can cause a fatal loss of the patient [5]. We need to worry about security or privacy protection when it comes to patient identification. To share and open the data we must provide a secure mechanism to guarantee data privacy and protection [6].

Semantics of data or semantic interoperability is the biggest challenge. Health information originates from several sources, and can have different meanings. In addition, the biomedical field develops several terminologies and ontologies [7]. An alternative is to rely on initiatives such as OBO Foundry, which is based on formal ontologies [1]. Formal ontologies can enable effective semantic interoperability due to the disambiguity in the definition of terms [1].

Conclusion

Some common challenges are both linkage data errors and privacy, but the most important issue is interoperability. Healthcare information comes from heterogeneous sources, and there are different terminologies in use.

Indeed, in recent years a proliferation of policies, standards and norms have been observed in many countries. The creation of these instruments follow a trend of modernization of public administration to provide better services to citizens. Overcoming these challenges, a global data openness in health is able to permit sharing, retrieval, combination and analysis of a multitude of health information.

Acknowledgments

We thank CNPq and CAPES for the granted scholarship in doctoral research for the first author.

References

[1] Smith et al. The OBO Foundry. Nature biotechnology 2007: 25 (n. 11): 1251-1255.

[2] Berners-Lee, T. Design: Linked data 2006.

[3] De Almeida, M.; De Mello, M. O uso da técnica de linkage de SIs em estudos de coorte sobre mortalidade neonatal. Rev. S. Púb. 1996: 30 (n. 2): 141-7, 1996.

[4] Campos, L. et al. Dados Abertos Interligados e o espaço do profissional de informação: XIII ENANCIB 2012. RJ.

[5] Harron, K., et al. Evaluating bias due to data linkage error in electronic healthcare records. BMC medical research methodology. 2014: 14 (n. 1): 36.

[6] Kelman, C. W. et al. Research use of linked health data: a best practice protocol. Australian and New Zealand journal of public health. 2006: 26(n. 3): 251-255.

[7] Word Health Organization. Management of patient information: trends and challenges in Member States: based on the findings of the second global survey on eHealth. Global Observatory for eHealth series. 2012: 6.

Address for correspondence

fernandafarinelli@eci.ufmg.br

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1123

Facilitating Full-text Access to Biomedical Literature Using Open Access Resources

Hongyu Kang[a], Zhen Hou[a], Jiao Li[a]

[a] *Institute of Medical Information &Library, Chinese Academy of Medical Sciences, Beijing, China*

Abstract

Open access (OA) resources and local libraries often have their own literature databases, especially in the field of biomedicine. We have developed a method of linking a local library to a biomedical OA resource facilitating researchers' full-text article access. The method uses a model based on vector space to measure similarities between two articles in local library and OA resources. The method achieved an F-score of 99.61%. This method of article linkage and mapping between local library and OA resources is available for use. Through this work, we have improved the full-text access of the biomedical OA resources.

Keywords:

Full-text Access; Open Access; Biomedical literature; Digital Library

Introduction

Open access (OA) articles in the biomedical field have a significant share. With the increasing volume of the full-text biomedical articles available online, researchers prefer to get them through internet rather than visiting a library. In this paper we describe and evaluate a mapping algorithm that matches articles stored in local libraries and biomedical OA resources. The algorithm uses the components of article title and publication year. We applied our method to map articles in NSTL (National Science and Technology Library), China, at http://www.nstl.gov.cn/ and PMC (PubMed Central), U.S., at http://www.ncbi.nlm.nih.gov/pmc/. We linked PMC's full-text articles to those at the NSTL based on the mapping results.

Methods

Our method improves full-text access by linking the articles in local library to an OA resource. It consists of a Vector Space matching algorithm that calculates similarity between two article titles. We define the article titles in local library and OA resource as vectors. Each dimension corresponds to a separate word. If a word occurs in the title its value in the vector is non-zero, else it is zero. We define the cosine of the angle between the vectors as the similarity between the two titles. A cosine value of one means the two title vectors match. We used precision, recall and F-score to optimize the mapping results. The workflow is shown in Figure 1.

Results

To evaluate the algorithm mentioned above, we included eight journals in the training set with 14267 articles in NSTL and 38815 in PMC. We used a critical threshold of 0.85 for F-score to evaluate this method. The precision and recall of the method were 100% and 99.22% respectively, resulting in an F-score of 99.61%. The result indicates that the number of the overlap between NSTL and PMC for year 2013 is 24,250, which represents 25.53% of the PMC dataset. The URL linking of the PMC full-text articles can also be given on the NSTL, based on the mapping results. Using this method, NSTL users can easily get an access to full paper by clicking the 'Full-text Reading' button.

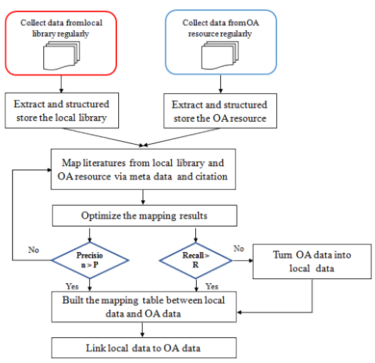


Figure1–The workflow of article resource matching

Conclusion

We have developed and implemented two components of a computational method that match the articles from different OA resources. The method successfully matched 25.53% of PMC articles to those in the NSTL records for 2013. The algorithm described in this paper was tested and verified on biomedical domain, using proper nouns in biology and chemistry. We believe the algorithm can also be applied to other sources of data with two modifications – by adjusting the critical threshold of the similarities according to the specific domain, and by adding the citation features besides tittle to the matching algorithm. It is clear that mapping the literature data from local library database and OA resources will facilitate linking citations to full text from a local library in an efficient manner. We plan to further explore the methods of deeper integration to make full use of OA resources.

Acknowledgement

This research is supported by the program of NSTL (National Science and Technology Library, China).

Address for correspondence

Jiao Li, Ph.D., Institute of Medical Information & Library, Chinese Academy of Medical Sciences, 3rd Yabao Road, Chaoyang District, Beijing100020, China. Email: li.jiao@imicams.ac.cn

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1124

What Medical Informaticians Do With and Think About an International Medical Informatics Listserv: Member Survey Preliminary Findings

Craig Kuziemsky[a], Martha B. Adams[b], Bonnie Kaplan[c], Kourosh Ravvaz[d], Ross Koppel[e]

[a] *Telfer School of Management, University of Ottawa, Ottawa, ON, Canada*
[b] *Division of General Internal Medicine, Department of Medicine; Division of Clinical Informatics, Department of Community and Family Medicine; Duke Center for Health Informatics, Duke University Medical Center, USA*
[c] *Yale Center for Medical Informatics at the Yale University School of Medicine; Yale Interdisciplinary Center for Bioethics; The Information Society Project at the Yale Law School, Yale University, USA*
[d] *Aurora Research Institute, Aurora Health Care, Milwaukee, WI, USA*
[e] *Sociology Department; Center for Clinical Epidemiology and Biostatistics, Medical School; LDI, Wharton School, University of Pennsylvania, Philadelphia, PA, USA*

Abstract

A survey of members of the American Medical Informatics Association (AMIA) listserv Forum on implementation and optimization asked how members perceived the Forum, and suggestions for improvement. Respondents appear to be remarkably engaged with the Forum's debates, information sharing, educational and practical teachings, comments, and immediacy.

Keywords:

Listserv; Informatics; Interaction; Sharing.

Introduction

In 2012, the American Medical Informatics Association (AMIA) created the "Implementation and Optimization Forum," (hereafter, the "Forum"). It quickly became a virtual meeting room for more than 1,200 medical informatics experts from all over the world. Discussions ranged from known implementation challenges to the current scientific literature. This poster extends our previous work on the types of content and discussions on the forum.[1]

Methods

We surveyed listserv members via SurveyMonkey to gain insight on their use of and perceptions about the listserv. We report preliminary results.

Results

Respondents appreciated the many insights posted on the listserv. Of N=173 respondents, 44% reported reading postings daily, 40% catch up on missed readings several times weekly, while 10% read the Forum at least weekly. Another question asked what attracts members to the forum. 172 respondents read it because they appreciate the interesting comments (66%), differing perspectives (63%), learning about important issues (61%), the range of experiences (58%), keeping abreast of developments (55%) and controversies (52%), and keeping connected. Respondents reflected on likes, dislikes, and suggestions for improvement (See Table 1).

Table 1– Sample comments about the forum

• What's so great is that the topics are spontaneous, and not the assignment of some publication/agency. Keep it up exactly as it is -- including the full thread of the conversation with each post (essential!). I must say that I'm surprised that there's not more controversy over the doom-and-gloom topics, but I'm also glad, because I usually agree completely with the savvy list pundits.
• It would be awesome to keep the topics more related to implementation, less in the "abstract" and more directly relevent to actual implementation logistics, costs, strategies etc.
• I thoroughly enjoy the listserv although some people do monopolize it occasionally.
• Would be great to make topics searchable if this list-serve were viewable elsewhere beyond the email stream. Inclusion of clinicians at key decision points in HIT design, implementation, AND POLICY/REGULATION. Addition of next/previous buttons to move directly back and forth between posts in a thread, although not attaching all the previous posts might make this unnecessary.
• I enjoy this listserv. In addition to the topics that are currently posted about most, which seem to be about system interoperability and safety and the lack thereof, I'd like it if more people felt comfortable posting about any informatics issue of potential interest. Let's encourage that.

Conclusion

The listserv offers a virtual global community that spans multiple disciplines and perspectives, and an underlying goal of improving healthcare IT. Respondents appear to be remarkably engaged and pleased with the Forum's debates, information sharing, educational and practical teachings, comments, and the sense of immediacy.

References

[1] Adams MB, Kaplan B, Sobko HJ, Kuziemsky C, Ravvaz K, Koppel R. Learning from Colleagues about Healthcare IT Implementation and Optimization: Lessons from a Medical Informatics Listserv. J. of Medical Systems 2015: 39 (1): 1-6.

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1125

A Comparison between LMS tools to support e-health educational activities

Magdala de Araújo Novaes[a], Gabriel Soares de Vasconcelos[ab], Jackson Raniel Florencio da Silva[ab]

[a] *Núcleo de Telessaúde, Hospital das Clínicas, Federal University of Pernambuco, Recife, PE, Brazil*
[b] *Center of Informatics, Federal University of Pernambuco, Recife, PE, Brazil*

Abstract

The objective of this study is to understand how a Learning Management System (LMS) plataform is used in a telehealth center to support two virtual learning enviroments focused on the education of the healthcare professionals and the students of a medical schoool. The study outcome is expected to provide indications towards choosing a better LSM for the telehelath center to support their educacional activities.

Keywords:

Learning management system; medical education; e-health.

Introduction

Learning Management Systems (LMS) provide tools and functions to support teaching and learning processes, and include course management tools, online group chat and discussion, homework collections and grading, and course evaluation [1][2]. Chosing an appropriate LMS is one of the main decisions before starting an e-learning project. LMS choice can be a decisive factor in the implementation and support of the project and involves administrative management, financial costs, and human resources [3]. During the last decade, many comparisons between LMS systems have been performed, which varied from simple comparisons – e.g., between few LMS features – to complex comparisons [4][5]. The aim of this study was to assess the open source LMSs, Moodle [6] and Sakai [7], to support e-health educacional activities by conducting literature review and practical evaluation [8].

Methods

The evaluation focused on the features of LMS and a combination of several frameworks. The top open-source LMS options provide: Feature-rich toolsets; Enterprise-grade stability, scalability, and security; a high degree of control and flexibility; and Generally lower long-term costs than commercial options. In addition, the LMS system should be able to accommodate and manage a huge amount of information required by the medical and educational content, including different subjects, lecture notes, lecture videos, text books, videos of surgeries, and radiology images [8].

Results

Choosing the right LMS for one's needs can be hard. Ramesh [9] has developed and presented a rubric to evaluate LMSs. He has also evaluated the LMSs Moodle and Sakai against the rubric, and arrived at a conclusion that Moodle is a better LMS to deploy for the chosen set of criteria.

According to [10], if customization, reporting and analytics, and collaboration are high priorities for e-learning initiative, Sakai is probably the best option. For the ease of use, extensibility, and a wide customer-base - as well as support and service vendors - Moodle may be ideal. Albarrak et.al. [5] indicated the maturity and advanced features in both Moodle and Sakai.

Conclusion

The choice is based on an organization's needs. With the given variability in functionality, cost, and hosting options, the key to successfully selecting and implementing an LMS is careful planning. It's critical to evaluate your organization's needs and requirements, as well as budget, staffing, and other available resources. Careful consideration of all these factors will help to choose the best LMS.

References

[1] Ceraulo, S, "Benefits of upgrading to an LMS". Distance Education Report 9, 9 (May 2005), 6–7.
[2] Yildirim, S., Temur, N., Kocaman, A. and Goktas, Y," What makes a good LMS: An analytical approach to assessment of LMSs." In Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training. (Istanbul, Turkey, May 31–June 2, 2004), 125–130.
[3] Akagi, A. "A escolha do LMS para projetos de EaD", website imasters.com.br
[4] Kljun, M., Vicic, J., Kavsek, B., Kavcic, A, "Evaluating Comparisons and Evaluations of Learning Management Systems", ITI
[5] Albarrak AI, "Designing E-learning System in Medical Education: A Case Study". International Journal of Excellence in Healthcare Management, 3 (1): 1-8, 2010.
[6] Moodle website www.moodle.org
[7] Sakai website www.sakaiproject.org
[8] Albarrak, A.I., Aboalsamh H. A. and Abouzahra, M, "Evaluating learning management systems for University medical education", Education and Management Technology (ICEMT), pp.672-677, 2010.
[9] Ramesh, V.M. ; Ramanathan, C., "A rubric to Evaluate Learning Management Systems",Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on, 73 – 77, 2013.
[10] Monarch Media, Inc., "Open-Source Learning Management Systems: Sakai and Moodle", Business White Paper, 2010.

Address for correspondence

gabriel.vasconcelos@.ufpe.br

This page intentionally left blank

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

Subject Index

3-D computer game 974
3D printed model 1025

A
AAC 589
access logs 3
access to information 153, 908
accessibility 3
active learning 35
acute healthcare delivery 912
adipose tissue 726
ADL 827
Adolescent 1099
adoption 17
adult (MeSH M01.060.116) 168
adverse drug events 950
"AEROS" 1068
Africa build project 1018
aged 964
aged care 910
aggregation 584
aging 462
air pollution maps 1048
alert fatigue 285
alert management engine 285
alert management strategy 285
algorithms 544, 1027
allergic reaction 320
allergy risk 1113
Alzheimer's disease 731
ambient assisted living 1060
ambulatory surgical procedures 894
Android 1024
angiotensin converting enzyme inhibitors 609
angiotensin II receptor blockers 609
antidepressants 867
app(s) 999, 1024
archetype(s) 64, 178, 207, 827, 881, 952, 1091
archetype development process 938
Arden syntax 955, 1111
Arden syntax rule engine 950
artificial intelligence 315, 1032, 1066, 1074, 1075
Asian American 996
assessment 902
assisted diagnosis 132
assistive technology 800
association analysis 1072
association learning 711
attitude to computers 264
audit trail 1080
automated procedure assessment 163
automated surveillance and monitoring 295
automatic audit system 926
automatic data processing 1056
automatic reasoning 564
awareness 103

B
bachelors programme 1016
barcoding 919
baseline patient model 549
bayes theorem 1075
Bayesian network 259
behavioral health 386
behavioral model 247
behavioural medicine 148
bibliographic 867
bibliometric mapping 1014
bibliometrics 1004
big data 69, 983
bigram filtering 1033
biochemistry 701
biological ontology(ies) 832, 1108
biomarkers 696
biomedical domain 815
biomedical informatics 609
biomedical literature 1123
biomedical ontologies 1092, 1095, 1103
biomedical research 913, 1053
biomedical technology assessment 482
biomedical technology intervention 148
biomedical terminology(ies) 716, 785
bipolar disorder 741
bleeding 721
blood transfusion 721
blood typing 559
bluetooth 198
board certification 501
bone tumor 672
BPMN 1118
Brazilian Portuguese 1022
breast cancer 746, 1104
breast cancer management 264
business value of IT 183

C
CAM-ICU 899
cancer 511, 1070
cancer diagnosis 882
cancer survivor 113, 142, 996
capacity building 520
cardiac arrest 974
cardiac fat 726
cardiac function 653
cardio pulmonary resucitation (CPR) 1062
cardiology 69
cardiovascular diseases 290
cardiovascular pregnancy complications 976
cardiovascular procedures 837
care coordination 21
case based reasoning 862
case study 17, 959
case-based reasoning 951
Casemix grouper 1090
CDA repository 173
CDS 270
CDSS 554, 955
certification process 227
cervical vertebrae 1038
chemotherapy protocols 1121
chest CT images 1046
chest wall allograft 1026
chest wall tumor 1026
children 554, 889
Chinese clinical text 624
Chinese hospital 207
Chinese language 979
choice architecture 148
chronic disease 487, 1039
chronic disease management 648
chronic gastrointestinal disorders 325
chronic heart failure 74
chronic illness 148, 842
citation ranking 1004
citizens 376
classification 721, 726, 746, 790, 941
classification system 247, 943
clinical 949
clinical alarms 918
clinical alerts 295
clinical and translational informatics 736
clinical application 198
clinical data modelling 1091

clinical data repository(ies) 207, 559
clinical decision support 947, 950, 1088
clinical decision support system(s) 93, 118, 259, 264, 275, 290, 300, 329, 419, 701, 857, 946, 948, 951, 956, 958, 1030, 1095, 1096, 1119, 1121
clinical drug hypersensitivity/ prevention & control 242
clinical guidelines 939
clinical informatics 338, 501, 901
clinical knowledge representation 1111
clinical models 958
clinical natural language processing 624
clinical notes 1033
clinical pathway(s) 682, 711, 1077
clinical phenotyping 1098
clinical practice guideline(s) 515, 952, 953, 1039
clinical prediction 696
clinical reference model 805
clinical research 448, 564, 924, 930
clinical research infrastructure 668
clinical task 1031
clinical trial(s) 534, 569, 930
clinical trials as topic 884, 1029
cloud computing 648, 691, 929
cluster analysis 579, 1072
clustering 746
coding system 1036
cohort analysis 1053
cohort study 876
cohorts 687
collaboration 103
colonoscopy 614
combination drug therapy 663
common data elements 1091
communication board 800
community care 429
community health planning 1009
community pharmacy 982
ComPacks 429
comparative effectiveness research 584
competence 993
complementary and alternative medicine 785
complex needs 963
complex systems 977
comprehensive health care 462
computational clusters 1082
computed radiography 914
computed tomography 726
computer 942
computer communication networks 1082
computer games 348, 363
computer interpretable clinical guidelines 980
computer literacy 971
computer system(s) 434, 897
computer systems evaluation 482, 969
computer-aided diagnosis 906
computer-assisted drug therapy 242
computer-interpretable clinical guidelines 954
computerized 30
computerized decision support 270
computerized medical records systems 242
computerized provider order entry 917, 940
computerized systems 183
concept dictionary 780
conceptual framework 857
conditional probability tables 259
confidentiality 3, 913, 919
Confocal 672
connected health 123
consensual reflex 906
consent 584, 984
consumer behaviour 108
consumer eHealth 968
consumer health informatics 333, 358, 391, 810, 973
consumer health information 84, 907, 1089
consumer informatics 338
consumer involvement 401
content-based classification 771
content-based image retrieval 1020
contingency plan 173
continuing medical education 372
continuity of care 936
continuity of patient care 935
control 242
controlled 821
controlled vocabulary 1089
COPD 910
cost-benefit 981
CPR 974
creatinine level 1063
critical care 920
cross sectional study 232
crowdsourcing 1033, 1091, 1113
cultural 434
curation workflow 1061
curriculum 525, 1016
CVD 1024
cyber security 915
cytochrome P-450 CYP2D6 946
cytogenetics 1037

D
data 529
data architecture 1055
data collection 45, 453, 584, 852, 994, 1056
data completeness 885
data curation 1056, 1107
data display 964
data exchange 1102
data fusion 731
data governance 7
data integration 682, 1019, 1049, 1051, 1061
data interoperability 1086
data linkage 584, 1052
data management 668
data mining 658, 726, 711, 1032, 1040, 1072, 1074, 1075, 1094
data preprocessing 1069, 1070
data quality 852, 882, 993, 994
data quality assessment 1069
data registration 997
data representation 1044
data reuse 997
data set generation 643
data sets 584
data sharing 1005
data staging 1055
data standardisation 1086
data storage and retrieval 529
data transformation 1055
data validation 882
data warehousing 1060
database(s) 574, 852, 867, 933, 1094
databases, factual 1101
dataset 1053
de-identification 1033
decision aid 905
decision fusion 300
decision making 1054, 1121, 1122
decision models 1070
decision support 1020
decision support system(s) 108, 242, 305, 564, 925, 949, 959
decision support techniques [l01.700.508.190] 280
decision-making system in medicine 957
deep learning 624
definitions 1092
deidentification 584
delirium 899
dental electronic health records 1081
dentistry 163
depression 629, 1099
depression relapse 741
dermoscopy 691
developing countries 520, 780, 883, 975, 1106

diabetes 1081, 1078
diabetes complications 961
diabetes mellitus 84, 93, 1072
diagnosis 995
diagnosis-related groups 315
diagnostic criteria 1093, 1097
dictation 878
digital image detection 1042
digital inclusion 970
digital library 1123
digital mammography 914
digital patient model 948
digital stethoscope 976
disaster 902
disaster response 1008
discharge summary 936, 1120
discourse-level analysis 539
discriminant power 1044
disease management 889
disease self-management 93
disseminated intravascular coagulation 956
dissemination 497
distance education 847, 1112
distant supervision 35
distributed decision support 891
doctor-patient relations 153
documentation 821, 945, 1116
downtime 472
driver genes 658
drug allergy 1088
drug interactions 949
drug labeling 1039
drug monitoring 950
drug repurposing 1037, 1051
drug repurposing
drug therapy 1039
drug utilization review 961
drug-related side effects and adverse reactions 242, 1030
DSL 955
DWH 955
dynamic time warping 163

E
e-Epidemiology 320
e-Government 12
Ebola 916
ECG 1044
economics 414
economics of health information systems 410
education 26, 506, 525, 1001, 1011, 1015, 1017
educational 434
educational models 901
efficiency 945, 981
eHealth 12, 79, 89, 153, 376, 391, 396, 434, 487, 648, 800, 876, 891, 898, 908, 937, 942, 971, 975, 981, 992, 1010, 1014, 1015, 1113, 1114, 1125
eHealth information systems 12
eHealth literacy 358
eHealth strategy 492
EHR databases 956
EHR documentation 255
EHR optimization 7
EHR use 255
elderly (MeSH M01.060.116.100) 168
elective surgery cancellations 438
electrocardiogram 98, 900
electrocardiography 69
electronic dental records 17
electronic health card 492
electronic health record(s) (EHR) 21, 26, 30, 40, 45, 50, 55, 64, 153, 158, 173, 232, 448, 584, 706, 721, 780, 881, 891, 884, 821, 875, 877, 878, 885, 903, 908, 931, 938, 946, 952, 961, 967, 997, 1009, 1030, 1050, 1055, 1063
electronic medical record(s) (EMR) 237,, 506, 639, 711, 880, 884, 879, 883, 904, 916, 932, 942, 966
electronic prescription 237, 982
electronic tablets 554
electronic whiteboards 477
element level 1109
eletrocardiography 989
email 401
embedded software certification 227
embedded systems 98
emergency department 875
emergency department design 960
emergency department information system 1001
emergency medicine 329
endocrinology 990
engineering [j01.293] 928
environmental factors 1048
environmental health research 1061
epicardial fat 726
epidemiology 766, 925
epilepsy 1116
equipment safety 127
ER 942
error evaluation 672
error sources 994
ESB 955
Essomenic 429
estimated glomerular filtration rate 696
ethics committees 579
Europe 1005
evaluation 497, 927, 968, 980, 1112
evaluation criteria 954
evaluation of learning 1112
evaluation studies 406, 917
event logs 310
event prediction 1065
evidence-based medicine 515
exercise 1024
exercise therapy 1095
expert and novice users 1023
expert system 259
eye tracking 84

F
factual 852
family history 604
feasibility studies 1029
federated query 1055
feedback 424
FHIR 932, 955
fibromuscular dysplasia 217
field programmable analog array 98
follow-up recommendation 1028
food allergy 320
foodborne diseases 766
formalization 1093
foster care 886
frames 815
framework 270
full-text access 1123
fuzzy methodologies 1111
fuzzy sets and fuzzy logic 295

G
games for health 386
general practitioner 376
genomic alterations 658
genomics 766, 1052
geo-referenced data 766, 1048
geographic information systems (GIS) 396, 1009
georeferencing 925
Ghana 1006
Google glass 901
governance 1091
government 983
grid node database 1055
group practice 852
group-based trajectory model 1063
GSM 396
guideline adherence 264
guidelines computer assisted-instruction (GCAI) 1062

H
h-index 1004
HACVT 1086
Haiti 883

haptics 163
Haralick features 1047
HCV 515
health 1018, 1024
health alerts and notifications 118
health care 212
health care economics and organizations [n03] 1105
health care management 953
health care quality 919
health care quality, access, and evaluation [n05] 1105
health consumers 810
health examination 975
health informatics 520, 891, 1002, 1059
health information 1109
health information exchange 21, 60, 217, 677, 931, 1007, 1009
health information management 978
health information seeking 996
health information system(s) (HIS) 30, 406, 892, 915, 935, 948, 967, 985, 993, 1000, 1012, 1100
health information system personnel 1000
health information technology 410, 1015
health information technology 401, 907, 962
health IT integration and application 1111
health literacy 358, 971, 1015
health outcomes research 410
health personnel 153, 997
health policy 429
health policy [I01.655.500.608.400] 1105
health promotion 380, 979
health record 886
health self-management 333
health services 343, 1108
health services administration [n04] 1105
health services needs and demand 1009
health services research 448, 574, 894, 977
health system strengthening 993
health telematics 492
health terminology standards 776
health workforce 677
healthcare 183, 222, 338, 1008, 1015, 1110, 1043
healthcare data mining 751
healthcare personnel 908
healthcare quality improvement 419
healthcare simulation 960
healthcare survey 153
healthcare system 921
healthcare-associated infections 295
healthy lifestyle 113
hearing aids (MeSH E07.814.458) 168
heart auscultation 976
heart diseases 343
heart diseases/rehabilitation 424
heart failure 40, 599, 609
hemoglobin (Hb) 975
herbal and nutritional supplements 785
heterogeneous 731
heuristic evaluation 358
Hidden Markov Model 756
high-frequent sentences 1033
hip replacement arthroplasty 1077
HIS 1068
histological techniques 672
HIV 467, 515, 677, 880
HL7 12, 932, 933, 942
HL7 FHIR 1060, 1098
HL7-CDA 285
home blood pressure 751
home care services 895
home telemonitoring 74
HONcode 1064
hospital 1106
hospital administration 944
hospital costs 315
hospital environment 310
hospital information system(s) 193, 198, 212, 242, 310, 482, 904, 917, 924
hospital IT systems 222
hospital network 915
hospitalization 60
hospitals 1010
HPV vaccines 761
HR. alert threshold 1041
HTML5 1078
human factors 338, 842
human resources 993
human-computer interaction 965
hyperglycemia 939
hyperkalaemia 949
hypertension 132, 810, 889
hypoglycemia 939
hypothesis generation 1113

I

ICD 790
ICD 9CM 1090
ICD-11 content model 1097
ICD10 1068
ICHI 1090
ICTs adoption 921
identification 919
IHE 12, 222, 933
IHE integrating the healthcare enterprise 911
image management system integration 930
image processing 896, 1042
image retrieval 1046, 1079
image segmentation 1045
image-based surrogates 930
imaging biomarkers 930
imaging network 923
immunosuppression 1119
implementation 487, 997
in-situ 338
inactivity 89
incidental findings 1027, 1028
index 1083
indexing 529
indigenous 343
indirect care 255
individual conditional expectation 1073
individualized care 977
individualized medicine 1052
inductive logic programming 741
infectious disease 766
influence diagram 290
influnza vaccine adverse event 1076
informatics 443, 1015, 1091, 1124
informatics workforce 501
information and knowlegde sharing 892
information dissemination 515
information extraction 35, 1030, 1031, 1038
information literacy 971
information management [L01.399] 928
information model 936
information needs 810
information processing 691
information retrieval 534, 1020, 1053
information retrieval system 924
information sharing 1054, 1102
information storage 1067
information storage and retrieval 30, 544, 691, 1039, 1084
information systems 183, 217, 677, 943
information systems management 1002
information techonology 183
information visualization 966
informed consent 897, 1007
injuries 329
innovation 1015
inpatient data 232
inpatients 919
insomnia 515

insulin dosing 93
integrated database 1057
integrated health care systems 1053, 1054
integrated systems 682
integration of academia and industry 410
intelligent computer-aided instruction (ICAI) 1062
intensive care unit 251, 941
interaction 1124
interactive case simulation tool 515
interdisciplinary 1015
interfaces 1049
international classification for nursing practice 776
international perspectives 907
internet 372
internet of things 891
interoperability 12, 222, 677, 776, 932
interprofessional communication 1059
interruptions 103
intervention 142, 940
iPad 554
IS success 927
ISMS 911
ISO 11179 Model 1097
ISO 13606 881
IT Management 472

J
JIA 554
JLAC10 1050

K
k-mer 1083
karyotype 1037
kidney transplantation 1119
kidney transplantation (MeSH E02.870.500) 1058
kinetic energy index 653
knowledge 525, 1038
knowledge acquisition 716, 1115
knowledge base(s) 275, 1057, 1067, 1107
knowledge database 815, 1121
knowledge discovery 1071
knowledge management 26, 202
knowledge representation 594, 1096, 1115
knowledgebases 663
KPI 960

L
lab test master 1050
laboratories 917
laryngeal pathologies 1047
latency 887
Latin America 372
Latinos 984
lead user method 237
learning curve 443
learning health system 984
learning management system 1125
lessons learnt 980
limited English proficiency 380
linear programming 315, 1066
linked data 815, 1122
linked health data 1019, 1122
listserv 1124
literature based discovery 539, 1094
liver cancer 414
liver disease 414
local features 1046
log management 1080
longitudinal analysis 696
low and middle income countries 973, 1006
low-income 981
low-resource settings 927
lung cancer 1020

M
m-health 98, 900, 1015
machine learning 40, 629, 639, 691, 721, 741, 761, 867, 926, 1064, 1073
machine translation 979
malaria disease 1006
malignant melanoma 663
managed care 969
mapping 1036
maturity assessment 921
MDER 1043
MedDRA 1076
mediastinal fat 726
medical audit 419
medical audit/standards 424
medical chart 926
medical CPS 549
medical decision support systems 862
medical device(s) 222, 227, 453
medical device software 353
medical education 386, 1125
medical entity recognition 643
medical error 202
medical forms 837
medical graduate 525
medical history taking/methods 821
medical informatics 26, 35, 50, 406, 434, 599, 852, 942, 976, 1007, 1011, 1017, 1074, 1075, 1082
medical informatics applications 948, 1095
medical informatics applications 30, 544
medical information system 957
medical language processing 879, 1030
medical record(s) systems 30, 997
medical records 1040, 1067, 1101
medical waveform 1043
medication adherence 60
medication error(s) 903, 940
medication safety 232
medication systems 1106
medication therapy management 275
MEDLINE 716
melanoma 691
mental health 123
MeSH 529, 937
metabolic syndrome 414
metadata repository 1098
metadata standards 668
metamodel 800
MFER 1043
mHealth 79, 113, 127, 358, 386, 396, 648, 898
mHealth for logistics 1006
microscopy 443, 672
MIMIC-II 956
mind map 1077
MML 881
mobile 902
mobile application(s) 89, 93, 127, 329, 847, 894, 897, 899, 903
mobile device 847
mobile health 118, 123, 338, 467, 896, 959
mobile health vans 909
mobile phone 338
model calibration 696
model of clinical processes 957
model-driven development 800
modelling language 800
models, computer 1101
models, decision support 1101
Moni 295
monitoring 497, 995
monitoring health insurance 193
MOOC 372
mortality 60
MTS-Manchester 942
multi center database 1120
multi-agent systems 305
multi-criteria decision analysis 905
multi-relational model 741
multi-skill tasks scheduling 305
multidimensional data 1079
multilingualism 1036
multimodal 731
multiple myeloma 951
musculoskeletal disorders 756

N
N-gram 1064
named entity recognition 624
NANDA 1059
NANDA-I 247
narrative medicine 511
national surgical quality improvement project 706
nationwide healthcare systems architecture 12
natural language generation 594
natural language processing (NLP) 579, 589, 599, 604, 609, 614, 619, 629, 634, 815, 979, 1022, 1027, 1028, 1031, 1032, 1035, 1040, 1059, 1067
natural language processing [l01.224.065.580] 280
near field communication 900
negation detection 634
negotiation 305
nephrology 372
network analysis 21
neural network 624
NIC 1059
NOC 1059
non-direct care 255
Norway 438
NoSQL Database 929
nurse 942
nurse calls 103
nursing 492, 776, 1114
nursing administration research 1013
nursing cost accouting 944
nursing diagnoses 247, 943
nursing informatics 251, 912, 941, 969, 1012, 1013, 1016
nursing informatics competencies 1012
nursing information system(s) 247, 943, 944
nursing process 941, 943
nursing workload 255
nutrition 876
nutritional therapy: machine learning 325

O
observation 574, 1118
observational studies 687
observatory 1114
obstetrics 64
older adults 970, 978
OMICS 668
oncological networks 953
oncology 1065
online health forums 137
online systems 153, 908
ontology(ies) 113, 202, 275, 534, 619, 790, 795, 827, 891, 958, 1051, 1070, 1099, 1104, 1110, 1115, 1122
ontology classification 1022
ontology repository 1103
OntoMama 1104
open access 1123
open source software 45, 920
openEHR 45, 178, 207, 881, 1091
openmrs 916
operative 821
optical character recognition 896
optical flow 653
optical navigation 1026
order sets 939
organizational change 1002
organizational implementation 477, 1001, 1002
osteoarthritis 905
osteoporosis 1075, 1082
outpatient 885
outpatient pharmacy 982
overcrowding 305
OWL 827, 1093

P
PACS 929, 1046
pain 554
pain documentation 805
palliative care 963
pancreatic cancer 604
paper, medinfo 2015 79
partial dependence plot 1073
participatory design 391
participatory health 333, 910, 977, 1015
pathological findings 634
pathology 443, 1070
patient activation 333
patient care 1027, 1028
patient care plan 945
patient care planning 895
patient care team 1100
patient care team 21
patient centered care 910
patient data management software 920
patient data modeling 178
patient data privacy 453
patient discharge 492, 935
patient education 386
patient engagement 113, 348, 363
patient facing technology 158
patient generated healthcare data 158
patient health information 1033
patient journey mapping 391
patient journey modelling 429
patient monitoring 132
patient out-of-pocket payments 193
patient participation 153, 893, 971
patient portals 978
patient recruitment system (PRS) 884
patient registry 1049
patient safety 55, 202, 903, 912, 939, 1008, 1017, 1116
patient selection 534, 569
patient triage systems 920
patient-centered care 462
patient-centered outcomes research 584
patient-specific modeling 1101
patients 907
PCR 1083
PDCA 911
peadiatrics 963
pediatric emergency department 305
pediatric-kidney transplantation 1049
peer-to-peer 1103
PEHR 884
people-centered e-health 977
perceptions 985
performance measurement 188
periodontal disease 1081
permanent education in health 847
person-centred health care 977
personal health information systems 74
personal health record(s) 320, 893, 896, 907
personal health system 74
personal health technologies 842
personal healthcare 751
personalization 756
Personalized medicine 325, 1054
personel mobility 1018
pervasive healthcare 132
PET 653
pharmacogenomics 946
pharmacovigilance 950
pharmacy productivity 982
pharmacy workflow 982
phenotype 795
phenotype detections 559
Philips Gemini TF 64 TOF 653
physical agents 305
physical and rehabilitation medicine 290
physical exercises 970
physician documentation 1059
physician's practice patterns [N04.590.748] 280
physician-patient relationship 506
physicians 501
pituitary adenoma 178
point-of-care systems 701
positron emission tomography 653
post coordination 795

post disaster 79
post-coordination validation 1087
post-experienced eye-tracked protocol 84
postoperative care 894
PowerScribe 922
ppatient safety event reporting 188
practice guideline [V02.515.500] 280
practice guidelines as topic 275
pre-operative planning 438
predictive modeling 1069
prenatal care 64
preoperative virtual planning 1026
preventive medicine 910, 975
primary health care 888
primary healthcare 69, 987, 989, 1019
primer 1083
privacy 467, 893
probabilistc networks 290
probabilistic modelling 259
PROBE 554
problem list 877
problem oriented medical record 877
procedures classifications 1090
process analysis 1118
process assessment (health care) 212, 1052
process mining 310
process model 310
process optimization 438
professional competence 1013
professional education 458
professional review organisations 458
professional-patient relations 153
prospective genome-cohort study 1057
prurigo nodularis 1042
psychiatric nursing 919
psychiatry 736
psychoeducation 999
psychological 424
public health 448, 986, 987, 983, 988, 991, 1005
public health informatics 79, 380, 979
public health practice 380
public health reporting 1070
public health surveillance 761, 766
public sector 1000
pulmonary nodule 1079
pupillometry 906

Q
qualitative research 443, 890, 985
quality and safety 821
quality assurance 837, 953
quality control 852, 922, 988
quality data model (QDM) 1097, 1098
quality improvement 55, 852
quality indicators 212, 251, 419, 1110
quality of life 511, 998
“Quartile” 1068
query tools 1023
question classification 810
questionnaires 1029

R
R 1065
radiography image interpretation 1045
radiology 1027, 1028
radiology report(s) 634, 922, 1085
radiotherapy 933
randomized controlled trial 949
randomized experiment 188
re-identification 584
readability 1109
readmission prediction 1063
real-world evidence 687
recall bias 320
recurrence plots 1047
REDCap 554
redesign 1002
reference standards 30
registries 1091
regulation 127
rehabilitation 290, 348, 363, 992
rehabilitation (MeSH E02.831.800) 168
rehabilitation of speech and language disorders 50
relation extraction 539
reminder systems 242
remote monitoring technologies 74
remote sensing technology 964
report errors 922
reporting templates 1085
reputation 137
research informatics 410, 1107
research infrastructure 1005
retail pharmacy 982
retrieval methods 1067
RFID 904
RFID technology 1008
rhinosinusitis 259
rich media 999
risk 857, 1032
risk assessment 1054
risk factors 462, 1032, 1081
risk management 353
risk prediction 1081
roaming 198
routine health information system (RHIS) 993, 1000
Rubidium-82 653
rule engine 947
rules 1093
rules based reasoning 862
rural health services 909
Rwanda 193
RxNorm 785

S
safety 237
safety certification 227
sarcoma [C04.557.450.795] 280
scalable 887
screen film 914
screening 614
screening method 914, 1068
secondary usage 1068
secondary use of data 913
section classification 35
security 467
segmentation 726
selection bias 569
self report 511
self-assessment 971
self-care 320, 889, 898
self-management 108, 118, 123, 142, 842,898
self-monitoring 898,1102
self-quantification 333
semantic groups 771
semantic interoperability 958, 1036
semantic interoperability 564
semantic medline 539
semantic resources 815
semantic search 1022
semantic web 285
semantics 217, 1067, 1071, 1091, 1094
sensor network 198
serious games 386
service oriented architecture 45
service quality 927
service support 985
service training 1112
service-oriented architecture 270
sexuality 1074
shared decision making 148
sharing 1124
shelter 902
shiny 1065
sickle cell disease 898
simulation 549, 1062
simulation-based training 506
skill retention 974
skull base navigation 1025
smart home 983
smart mobiles 999
smart phones 127
smartphone application 123
smoking 1081
SNOMED CT 790, 795, 832, 1087
SOA 270

social media 137, 643, 761, 972, 1099
social network service 511
software (MeSH L01.224.900) 168
software agents 305
software design 368
sonification 1041, 1044
South Africa 1000
spanning tree 1077
SPARQL 815
speech impairment 800
speech recognition 922
speech therapy 986
spinal cord injuries 998
ST depression 1041
stakeholders 1000
standardization 222, 353, 477, 936, 1087
standards 529, 790, 932, 1102
state medicine 938
statistical analysis 108, 549
STD 515
stitching 672
strandards 933
stroke 348, 363, 992
structured data elements 7
structured EHR 878
summarization 594
supervised learning 706
support systems 511
surgical procedures 821
surgical site infection 706
survey 472
survival score 40
SWOT analysis 472
SWRL 1093
system 942
system architecture for nationwide ehealth 12
systems integration (H01.770.787) 1058
systems medicine 951

T
target audience 1109
task analysis 842
taxonomy 736
teaching 525, 1010
teaching materials 976
technical standards 26
technology acceptance 919, 962
technology evaluation 406
tele-homecare 648
tele-vital signs 648
telecardiology 991
telecare 891
teleconsultation 79, 888, 988
telehealth 888, 890, 891, 892, 963, 987, 990, 995, 1010, 1108
telemedicine 69, 497, 890, 894, 898, 937, 971, 986, 988, 989, 990, 991, 1010
telementoring 887
telemonitoring 89, 889
teleradiology 923
temporal abstraction 594
term occurrences 1085
terminology(ies) 26, 619, 780, 790, 941, 958, 1067, 1088
terminology as topic 1036
terminology fingerprinting 771
testing 549
text classification 629, 634
text mining 761, 882, 1037, 1120
text prediction 188
therapy adherence 682
thermal signals 746
time and motion study 255
time-series data analysis 751
tohoku medical megabank project 1057
tomography 1045
training 520, 1001, 1015
training (MeSH SP4.006.047.453.584) 168
transdisciplinary research 977
translation 380
translational medical research 1052
translational research 687, 1071
treatment 995
treatment decision 259
treatment effect 721
treatment recommendation 300
TRIAGE 942
trust 137, 467
TURF 965
twitter messaging 761, 766, 972
type 1 diabetes mellitus 972
type 2 diabetes 60, 682, 1048

U
UMLS 771, 785, 837
uncertainty 857
unified theory of acceptance 992
universal health coverage 193
unstructured information management architecture 604
unsupervised learning 736
update weighing attributes 1020
usability 84, 188, 338, 353, 358, 892, 965, 967, 968
use error 353
use of technology 992
user center design 967
user centered design 1006
user interface design 237
user-centered design 368, 386
user-centred design 892, 912
user-computer interface 368
USSD 396
utility theory 998
UX 966

V
VAERS 1076
validation 64
validity indexes 746
ventricular ejection fraction 599
vibrotactile biofeedback 756
viral load 880
virtual preoperative planning 1025
virtual reality 163
visualization(s) 1065, 1080
vital signs 918, 1035
vocabulary 217, 821
vocabulary, controlled 1084
vulnerability 462

W
wavelet transform 1047
wearable devices 898
web 142
web interface 589
web questionnaires 908
web service 1113
web tool 259
web- based communication 438
web-based reporting 320
web-based system 876
webapp 1078
WebRTC 887
wellness programs 487
Wilcoxon signed-rank test 1085
women 975
work sampling study 255
workflow 1100
workflow integration 930

X
XY-table 672

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

Author Index

A
Abbud-Filho M. 372
Abdaoui A. 137
Abedini M. 691
Abidi S.R. 118, 148, 285
Abidi S.S.R. 118, 285
Abin J. 12
Abramovicz-Finkelsztain R. 17
Abrams Z.B. 1037
Abreu Maia T. 938
Abu-Hanna A. 997
Acosta L. 923
Adam T.J. 785, 821
Adams M.B. 1124
Adlassnig K.-P. 295, 950, 1111
Afonso R.A. 983
Agbakoba R. 487
Aguiar Roza B. 1058
Ahmad A.M. 285
Ahmed A. 79, 975
Ahmed Z. 1090
Ahn S.-M. 882
Ainsworth J. 123
Akbarov A. 696
Ålander T. 153, 908
Alassia L.N. 3
Alavarce D.C. 1112
Alba P.R. 1035
Alemrares A. 967
Aljunid S.M. 1090
Alkmim M.B.M. 69, 986, 987, 988, 989, 991
Almalki M. 333
Almeida H.O. 549, 862
Alvarez M. 903
Álvaro A. 983
Alves C.S. 1104
Alzheimer's Disease Neuroimaging Initiative 731
Amaral H. 1007
Amith M. 113
Amoroso C. 880
Anand V. 554
Ancker J. 1023
Ando Y. 911, 933
Andor M. 1041, 1044
Andreão R.V. 888
Angulo F. 584
Anne A. 1018
Aonghusa P.M. 462
Aparecida Manenti S. 1074
Aponte-Tinao L.A. 672, 1026
Applegate E. 123
Araújo M.M. 976
Arachi D. 761, 867
Araki K. 1050
Aramaki E. 511, 1113
Ares F. 903
Arispe B. 921
Arlt F. 1030
Armstrong K. 894
Arnolds I.V. 1066
Arrais Porto M. 1046
Årsand E. 84
Arsoniadis E.G. 158, 706, 821
Asai T. 926
Assélé-Kama A. 217
Atique S. 914
Atwoli L. 520
Ayaya S. 1113
Ayuo P.O. 520
Azé J. 137

B
Baesse D.C.L. 847
Bagayoko C.O. 1018
Bahroos N. 1059
Baja C. 883
Bakken S. 467, 984, 1006
Bakker T. 1052
Balas E.A. 448
Balasubramanian A. 1059
Ball E. 916
Banerjee A. 947
Banerjee N. 947
Bannenberg U. 909
Baptista R.S. 1010
Barbieri A. 937
Barbosa Neto J. 970
Barbosa L. 589
Barbosa S.F.F. 251
Barcellos Almeida M. 1092, 1115
Barcellos de Almeida M. 1122
Barra C.M. 1022
Barsottini C.G.N. 17
Bartholmai B.J. 922
Bartz C.C. 1114
Bastião Silva L.A. 687
Bastien J.-M.C. 353
Bates D.W. 242
Baughman A.W. 629
Baum A. 386, 985
Baum A.J. 232
Beattie J. 529
Becker Sander G. 1007
Bedra M. 89, 889
Beeler P.E. 949
Beesley C. 604
Bellazzi R. 682, 1048
Bellika J.G. 887, 958
Bellwood P. 237
Ben-Assuli O. 1063
Benevenuto de Campos Lima C. 937
Benigno M. 1056
Benítez S. 3, 386, 877, 890, 942, 985
Benítez S.E. 232, 255, 472
Bergh B. 222, 884
Bernhardt-Melischnig J. 716
Bernth Ahrenkiel S. 497
Bertele P. 1005
Bertelsen P. 376
Bertoia A. 985
Bertulli R. 280
Besana P. 564
Best L. 1090
Betts H. 525
Beuscart R. 69
Beuscart-Zéphir M.C. 353
Bezerra J. 862
Bhuiyan Masud J.H. 434
Bielinski S.J. 946
Bigaton E. 1082
Bignens S. 1011
Bijakowski K. 325
Binotto A.P.D. 648
Biondich P. 932
Bisognin Ceretta L. 1074, 1075
Bittar A. 1067
Blair D. 946
Blasca W.Q. 168
Blaser J. 949
Blaszka-Jaulerry B. 264
Blažun H. 1014
Bleimeyer R. 946
Boccio C. 1025
Bock C. 964
Böckmann B. 953
Bodenreider O. 771, 795
Bolin P. 391
Bolle S.R. 438
Bonisson L. 991

Borbolla D. 877
Borhanian E. 569
Borycki E.M. 338, 506, 1016
Bosomprah S. 1018
Botti M. 912
Bouamrane M.-M. 487
Bouaud J. 264, 275
Bourgeois F.T. 867
Bouvry C. 1067
Bouzillé G. 564, 913
Bowen R.A. 123
Bowles K.H. 406
Bowman M. 1063
Boyd A.D. 1059
Boyer C. 1064
Boyle D.I.R. 1019
Bradford D. 343
Braga D.S. 1104
Brand Persson K. 790
Bras Da Costa S. 353
Brasil L.M. 1104
Bray B.E. 609
Brear H. 790
Bright A. 883
Bringay S. 137
Brito Alves de Lima G. 937
Brown B. 419, 696, 701
Brown W. 467, 584
Bruflat J. 946
Bruland P. 1042
Bucciolli Guernelli M. 1007
Buchan I. 123, 320, 419, 696, 701, 1005
Budrionis A. 887
Burgun A. 1005
Bürkle T. 1011
Burton M. 1059

C
Cabral Robinson C. 290
Cahan A. 1056
Cai Y. 1076
Cambon-Thomsen A. 1005
Cameselle L. 985
Camilo E.N.R. 906
Campbell J. 790
Campillo-Gimenez B. 564, 913
Campos da Costa L. 983
Campos A. 1108
Campos F. 173, 368
Cancio A.H. 472
Cândida Castro T. 1012
Cantarutti F. 1007
Cao M. 965
Capurro D. 380, 875, 885
Carøe P. 910, 1015
Caraballo P.J. 946
Carballo-Dieguez A. 467
Cardoso Coelho K. 1115
Cardoso Franco Ortiz D. 247
Cardoso Garcia V. 983
Cargnin D.J. 348, 363
Carlo L. 1006
Carmeli B. 280
Caroline F. 1018
Caron A. 69
Carrasco C. 1001
Carrier M. 1100
Carter P. 968
Carvalho B.C. 989
Casagrande R.A. 1082
Casali P.G. 280
Case J.T. 795
Casper G.R. 842
Cassettari Junior J.M. 1074
Castaño J. 634
Casteli C.P.M. 1013
Castrillón E. 1062
Castro Júnior E.F. 847
Catalán S. 875, 885
Çatalyürek Ü. 766
Cerra C. 1048
Cervi Prado A.L. 348, 363
Cesário Times V. 64
Cesar Gadelha Vieira A. 1007
Cesconetto S. 1074
Céspedes J. 132
Cha E. 939
Chae Y. 879
Chakravorty R. 691
Chalopin C. 30, 178
Chang F. 242, 629
Chang H.-F. 919
Chang P. 899, 974, 1024
Chao L. 736
Charles A. 920
Charretk N. 920
Chattopadhyay S. 947
Chatzimina M. 1117
Chaudhuri S. 380
Chazard E. 69
Chella M.T. 589
Chen E.S. 821
Chen F.-C. 919
Chen Q. 643
Chen Y. 965, 1076
Chêne G. 1005
Chenggang Z. 1083
Cheung N.T. 945, 1086
Chiang I.-J. 914
Chiovato L. 682
Choi C. 924
Choi Y. 453
Chute C.G. 790, 1093, 1097, 1098
Cimino J.J. 559, 579, 956
Cintho L.M.M. 952
Ciqueto Peres H.H. 247, 1112
Ciriaco Pereira D.L.N. 1095
Coenen A. 776, 1114
Coiera E. 761
Cojean-Zelek I. 264
Colburn D. 7
Coleman K. 912
Collins S.A. 7, 805, 907
Comunello E. 1074, 1082
Concannon T. 584
Concha P. 960
Conci A. 726, 746
Cook C.B. 93
Cordeiro d'Ornellas M. 348, 363, 1045, 1046
Cordeiro de Morais F.F. 726
Cording A. 458
Cornet R. 997
Corradi S.L. 888
Correa-Arango A.L. 1062
Corrêa Coronel C. 290
Correia S.E.N. 1047
Côrtes de Mattos Garcia M. 1074
Costa D. 1007
Costa H.D.R. 1104
Costa R.M. 906
Costa S.L.N.C. 1047
Costa Oliveira M. 929, 1020, 1079
Cotik V. 634
Couch P. 320
Couto Carvalho Barra D. 941
Coyte P. 894
Crooks J. 946
Crowner C. 619
Cuggia M. 564, 913
Cui Y. 974
Cummings E. 1016
Cunningham J. 1005
Curry J. 429
Cuzin-Kihl A.K. 1018
Cypko M.A. 259

D
D'Souza J.M.J. 962
Dagliati A. 682, 1048
Dai T. 810
Dailey M.N. 163
Dalevi D. 1065
Damasceno de Souza A. 1092
Darade P. 1081
Darmoni S.J. 212, 544, 1036, 1067
Davies I. 237, 928
Davies L. 123
Davis M. 691
Davoody N. 992
da Rosa M.I. 1075, 1082
da Silva Carlessi L. 1082
da Silva Dias R. 741
da Silva Klahr P. 290

da Silva I.T. 876
de A. Costa W.C. 1047
de Almeida Lopes Monteiro da Cruz D. 247, 943
de Almeida Pititto B. 876
de Araújo Brandão Couto E. 986
de Araújo Novaes M. 64, 1125
de Araújo J.M.F.R. 857
de Britto F.A. 895
de Bruijne M. 997
de Bruin J.S. 295, 1119
de Cassia Gengo e Silva R. 247, 943
De Cata P. 682
de Keizer N.F. 424, 997
de Lima Lopes P.R. 1010
de Lima Verde Brito T.D. 1010
de Souza A.C. 935, 936
Delamarre D. 913
Delaney B. 35
Deleris L.A. 462, 1032
Della Méa Plentz R. 290
Della Mea V. 790
Delta S. 1055
Demiris G. 964, 978
Dencker Johansen M. 1078
Denecke K. 30, 178, 259, 948, 1030, 1038
Deng Y. 1038
Denny J.C. 1098
Dérilus F. 883
Desai L. 979
Deserno T.M. 930
Deserno V. 930
Dew K. 979
Dexter P. 604
Dhombres F. 795
Dhopeshwarkar N. 242
Di Lella F. 1025
DiEugenio B. 1059
Diallo G. 1103
Días C. 687
Dias Flores C. 290
Dias da Silva T. 937
Didonet Del Fabro M. 827
Dillon T. 969
Dinevski D. 1094
Ding S. 974
Diodati G. 896
Divita G. 639
Dixon B.E. 1009, 1070
do Nascimento Givigi R.C. 589
do Prado Fay J.H. 1007
Döhler B. 1049
Dolamic L. 1064
Domb S. 506
Dongsheng Z. 1083
Dornelles Picon P. 1007
dos S. Diogo R.C. 247, 943
dos Passos M.G. 1075
dos Santos Brito K. 983
dos Santos Ramos P. 937
Dota E. 1007
Dourado E. 903
Dowie J. 905
Downing N.L. 931
Drews M. 909
Du J. 1076
DuVall S.L. 614, 1035
Duan H. 207, 658, 1031
Duda S. 883
Duda S.N. 994
Dufour J.-C. 564, 1005
Dugas M. 837, 1029
Duke J.D. 574
Dukuze A. 880
Dumesnil C. 69
Dumontier M. 1039
Dunn Lopez K. 1059
Dunn A.G. 761, 867
Dunn-Galvin A. 320
Duperval J. 883
Duplan M. 883
Durieux A. 264
Dutra I. 741
Duvauferier R. 564
Dyb K. 438

E

Edema N. 920
Egbert N. 492
Ehrler F. 898
Eisenberg M. 931
Elbakkoush A.A. 914
Elkin P.L. 619
Ellingsen G. 878, 934
Elliott M. 946
Endler G. 852
Engelmann U. 893
Erbs J. 1058
Esamai F. 520
Eschmann E. 949
Esteves Perche M. 1007
Evans R.S. 270

F

Fábio Maciel R. 1058
Farfalli G.L. 672, 1026
Farinelli F. 1122
Farkash A. 280
Farri O. 1022
Fatima Marin H. 17
Faysel M.A. 915
Feather J. 968
Fehre K. 950, 1111
Feijó Ortolani C.L. 1058
Fenske M. 492
Fensli R. 891, 892
Fernández M.T. 472
Fernández-Cean J. 372
Fernandes De Muylder C. 938
Ferreira E.B. 847
Ferreira S.R.G. 876
Ferris T.A. 1084
Fialek S. 1012
Ficheur G. 69
Fidalgo R.N. 800
Figueiredo Damásio J. 227
Filippo D. 634
Finger M. 1040
Finkelstein J. 89, 108, 889, 939
Fischer J. 584
Fischer T. 909
Fizzotti G. 998
Flatley Brennan P. 842
Florencio da Silva J.R. 1125
Florez-Arango J.F. 1062
Florian R. 35
Folchetti L.D. 876
Föller-Nord M. 973
Fong N. 506
Fonseca J.M. 290
Fontes C.A.P. 746
Forbush T.B. 614
Forcella J. 921
Forte Lombardi A. 1007
Fraccaro P. 701
França A.L.N. 989
França R.M. 847
Franagan B. 1077
Franco M. 877
Franco N.M. 800
Franklin A. 999
Fraser H. 916
Frederickson Comer K. 1009
Freire C.M.V. 976
Frey A. 1072
Friedmann I. 12
Fritz F. 927, 981, 1029, 1042
Fujita K. 966
Fung V. 1086
Furner B. 584
Furstrand D. 971

G

Gabarron E. 972
Gabrielson D. 946
Gadd C.S. 994
Gaebel J. 1030
Gaff C. 1052
Gaju E. 880
Galopin A. 275
Gamba P. 1048
Gang H. 906
Ganslandt T. 1005

Ganzinger M. 951
García S. 372
García G.M. 255, 890
García-Duque J. 50
Garcia Aurelio M. 967
Garcia D. 952, 1075
Garcia G. 890, 985
Garde C. 564
Garnavi R. 691
Gartner D. 315, 1066
Garvin J. 599, 609
Gascon G.M. 414
Gasenzer R. 1011
Gaspar J.S. 935
Gatewood L.C. 529
Gayraud O. 920
Ge C. 1031
Gen S. 1120
Geng Y. 113
George S.M. 996
Georgiou A. 917
Gerdes M. 891, 892
Gesner E. 7, 805
Ghosh P. 975
Giacaman P. 875
Giavarini A. 217
Gibaud I. 564
Gibrat J.-F. 1005
Gibson C. 766
Gibson P.J. 1009
Gicquel Q. 1067
Gil J. 921
Giunta D. 386, 985
Giunti G. 386
Giussi Bordoni M.V. 877, 890
Gjerstorff M.L. 1005
Goel S. 584
Goel V. 918
Goldbraich E. 280
Golden S.H. 939
Goldstein A. 594
Goldstein M.K. 599
Gómez A. 386, 896
Gómez A.R. 472
Gomide L.B. 1104
Gonçalves P.B. 940
Gonçalves V.H.L. 1104
Goncalves Barbosa L.M. 940
Gong Y. 188, 202
González Bernaldo de Quirós F. 3, 173, 232, 386, 434, 890, 896, 985, 1025
González Z.A. 255, 967
Gonzalez-Martinez F. 372
Gordon L.T. 448
Gorgônio K.C. 549
Goss B.C. 922
Goss F. 242
Grabar N. 815, 1064
Graffi Moltrasio L. 903
Grain H. 26, 458
Grando A. 93
Granja C. 438
Grannis S.J. 1070
Gray K. 333
Greaves W.W. 501
Greloni G.C. 372
Grieve G. 932
Griffith J. 338, 358
Griffon N. 212
Groat D. 93
Groll M.J. 1038
Grosjean J. 1036
Gude W.T. 424
Guillen S. 942
Guliev Y.I. 957
Gundlapalli A.V. (Adi V.) 501
Gundlapalli A.V. (Aditya V.) 501
Guo H. 810
Guo M. 974
Gupta N. 880
Gurr G. 21
Gusmão C. 959, 970
Gutierrez M.A. 653

H

Haak D. 930
Haatainen K. 1116
Haddad A.E. 1010
Haddad P. 183
Haddawy P. 163
Hägglund M. 391, 992
Hamai S. 1077
Hamilton A. 584
Hammadi S. 305
Hamon T. 815
Hamzah Aljunid S.M. 1090
Han B. 643
Hanmer L.A. 993, 1000
Hänse W. 1042
Hansen D. 343
Hardiker N.R. 776, 1114
Hartvigsen G. 84, 438, 887
Hartzler A. 978
Harzheim E. 990
Hasan S.A. 1022
Hasvold P. 887
Haux C. 951
Haux R. 909, 1060
He S. 21
He Z. 569
Heath J. 453
Heavirland J. 609
Heerema N.A. 1037
Heffer L. 1052
Heiden K. 953
Hejlesen O.K. 84, 1078
Hellwig T. 909
Hideto Y. 1120
Higgins T. 467
Hirayama Y. 1068
Hirokawa S. 1077
Hirsch D. 259
Ho K.H. 945
Höcker B. 1049
Hoerbst A. 1118
Hofer-Dückelmann C. 950
Hofmann S. 909
Höft-Budde P. 909
Hohmann S.F. 584
Holanda do Nascimento C. 983
Hollis S. 1065
Holm J. 1011
Holthe H. 74
Hong N. 1093
Hong S.K. 904
Hong Y. 1085
Hoshino A. 1068
Hoshino T. 1068
Hota B. 584
Hou L. 1061, 1109
Hou Z. 1109, 1123
Hourihane J. 320
Hovenga E. 26
Howe S. 270
Hripcsak G. 574
Hristovski D. 1094
Hruby G.W. 1023
Hu M. 79
Hu Z. 706
Hua L. 188
Huang D. 964
Huang P. 1056
Huang Z. 658, 1031
Hübner U. 492, 1072
Hulbæk L. 497
Hultman G. 158
Hunter D. 905
Hunter I. 962
Huser V. 574
Hwang M.-A. 897
Hynes D. 584

I

Iaione F. 98, 900
Indarte S. 434
Ingelbeen B. 1018
Iochpe C. 648
Ishihara K. 955
Ishii T. 902
Ishima M. 944
Islam Maruf R. 975
Issom D.-Z. 898
Ito M. 1095
Iwamoto Y. 1077
Iwashyna T.J. 1035
Iyengar S. 999

J
Jackson M. 1056
Jafarpour B. 285
Jäger R. 909
Jahansouz C. 158
Jang B.-H. 879
Janies D. 766
Jansen K. 776
Jansen L. 1060
Jaulent M.-C. 217
Jazayeri D. 916
Jenders R.A. 1055, 1096
Jensen R. 1013
Jeon E. 1099
Jeong I.c. 108
Jeunemaître X. 217
Ji M. 899
Jia Z. 1031
Jiang G. 1093, 1097, 1098
Jiang M. 624
Jimeno-Yepes A. 643
Jochim C. 1032
Joe J. 978
Johansen M.A. 878
Johnson K.V. 270
Johnstone A. 1052
Jones Ferreira de Lucena D. 1020
Jones M. 1035
Jonnagaddala J. 1069
Jouini W. 564
Joukes E. 997
Joulakian M.B. 212
Ju M. 1031
Jung H. 1099
Jung W.-M. 879
Júnior C.A.E.M. 589

K
Kabino K. 930
Kagawa Singer M. 996
Kakkanatt C. 1056
Kälviäinen R. 1116
Kamada T. 911, 933
Kang H. 1123
Kanter A.S. 916, 1006, 1088, 1106
Kaplan B. 1124
Kapur S. 123
Karara G. 193, 482
Karema N. 880
Karmen C. 951
Karunanithi M. 343
Kasthurirathne S.N. 932, 1070
Kastrin A. 1094
Katsuhiko T. 1120
Katz N. 990
Kaufman D. 93
Kay M.S. 98, 900
Kayiganwa M. 880
Kayser L. 971
Kazantzaki E. 1117
Kebede M. 981
Keenan G.M. 1059
Keny A. 780
Kergourlay I. 1067
Kesler D. 501
Kesterson J. 604
Kezadri-Hamiaz M. 1100
Khairat S. 158
Khairat S.S. 401, 529
Khalifa A. 1018
Kho A. 584
Khodambashi S. 954, 980
Kiefer R. 1098
Kim A.R. 142
Kim D. 1043
Kim J. 1043
Kim S. 1043, 1090
Kim S.I. 1011
Kim T.Y. 776
Kim W.-S. 882
Kim Y. 599, 609
Kimura E. 955
Kimura M. 933, 1072
Kinnunen U.-M. 1116
Kipps A. 918
Kirchhoff K. 979
Kiriinya R. 677
Kite B. 414, 1107
Kivekäs E. 1116
Klein G.O. 901
Klemets J. 103
Knaup P. 951, 1049
Kobayashi R.M. 1013
Kobayashi S. 881, 1050
Koch L. 259
Koch S. 391, 898
Kodama N. 751
Kodie R. 1006
Kohl C.D. 1049
Kokol P. 1014
Koldkjær Sølling I. 910, 1015
Koller W. 295
Kolter T. 1030
Komandur Elayavilli R. 539, 1033
Kondylakis H. 1117
Koppel R. 1124
Kor D. 721
Köster L. 1049
Kothari C.R. 1071
Koumakis L. 1117
Koziol-McLain J. 968
Kraft T. 766
Kreuzthaler M. 716
Krishnan A. 604
Krive J. 410
Kropf S. 30, 178
Krupka K. 1049
Kuchinke W. 1005
Kuhlisch R. 492
Kukafka R. 108
Kullo I. 946
Kumagaya S. 1113
Kumar Meher S. 434
Kumar V. 434
Kume N. 881, 1050
Kuperman G.J. 907
Kuroda T. 198, 926, 966
Kury F.S.P. 579, 956
Kushniruk A. 506
Kushniruk A.W. 338, 358
Kutun T. 973
Kuwata S. 926
Kuziemsky C. 1100, 1124
Kwaan M.R. 158, 706, 821
Kwak I.J. 897
Kwaro D. 780

L
La Rosa F. 942
Lafer B. 741
Lähteenoja M. 982
Lai K. 242
Lai K.H. 629
Landoni M.C. 877
Landsberg G.A.P. 895
Lane S. 931
Lanyo K. 677
Larrea Armenteros O.U. 310
Larsen L.B. 84
Lasierra N. 1118
Lassoued Y. 1032
Lau M. 1086
Le T. 771, 964
Le X.H. 515
Lebech Cichosz S. 1078
Ledieu T. 913
Lee J.-H. 882, 924
Lee K. 1043
Lee Y. 882, 1043
Lefranc J.-P. 264
Lehmann C.U. 501
Lehmann M. 1011
Lei J. 624
Lei V.J. 629
Leite M.M.J. 1013
Leivas J. 925
Lelong R. 544
Lenz R. 852
Leong T.-Y. 731
Lepage E. 212
Leporati P. 682
Leshno M. 1063
Lettieri E. 1005
Leung T.I. 1039
Levy S. 963
Lewis S. 123

Li D. 539, 604, 1033
Li H. 658, 919, 1031
Li J. 810, 974, 1061, 1109, 1123
Li L. 917
Li S. 974
Li X. 89, 300, 711
Li Y. 1033
Li Y.-C. 574
Liang C. 202, 1051
Liaw S.-T. 458, 1069
Lima C.S.P. 251
Lima J. 970
Lima M.A.M. 549
Lin J. 1004
Lin L. 978
Lindgren K. 910, 1015
Lindner D. 178
Linhares de Souza Y. 1122
Liu H. 300, 539, 604, 711, 1033
Liu J. 1022, 1101
Liua H.-C. 919
Ljosland Bakke S. 1091
Lloyd B. 93
Lloyd J.F. 270
Löhnhardt B. 668
Longhurst C. 918, 931
Lopes da Rosa T. 937
Lopetegui M. 1001, 1087
Lopez A. 372
Lopez-Campos G. 1102
López-Nores M. 50
Lorenz S. 909
Lorier L. 372
Lovick S. 1065
Lovis C. 898
Lozano J. 921
Lu M.-R. 919
Lu X. 207
Luan Z. 969
Lucena A.M. 989
Luna D. 173, 368, 386, 434, 877, 890, 896, 967
Luna D.R. 232, 255, 472
Luna-Gomez I.F. 1062
Luque A.E. 515

M

MacKinlay A. 643
Macciocca I. 1052
Machado Ribeiro I. 1007
Machado E. 988
Machin M. 123
Madanian S. 1008
Madde N. 721
Madigan D. 574
Madsen I. 1016
Magrabi F. 448
Maia J.X. 988
Maia M.F.L. 847
Maia T.A. 935
Mair F. 487
Makhlysheva A. 972
Malykh V.L. 957
Mamas M. 696
Mambone Z. 1103
Mamlin B. 932
Mamykina L. 1006
Mancino A. 1025
Mancino A.V. 672, 1026
Mandirola Brieux H.F. 434, 942
Mankovich G. 1027, 1028
Mann H. 1065
Manohar N. 785
Mantovani Faustino de Carvalho D.C. 940
Manzi A. 880
Mar P. 7
Marc D.T. 529
Marcelin A. 883
Marchand G. 127
Marco L. 972
Marco-Ruiz L. 958
Marcolino M.S. 69, 986, 987, 988, 989, 991
Marcon Dal Sasso G.T. 941
Margolis A. 372
Marias K. 1117
Marino B.C.A. 69
Marinoni A. 1048
Marschollek M. 1060
Martínez E. 921
Martínez M. 896
Martínez-Costa C. 716, 832
Martin N. 979
Martin-Sanchez F. v, 333, 1052, 1102
Martinez S. 892
Martinot A. 305
Martins P.J. 1075, 1082
Martins T.B. 895
Marvin A.G. 842
Masayuki H. 1120
Masella C. 1005
Massari P. 212
Massaut J. 920
Massicano F. 534
Mathiesen K.S. 910
Matsubara N. 1068
Mauro A. 1001, 1087
Maviglia S. 7
Mayan III J.C. 232
Mazzoncini de Azevedo-Marques P. 1079
Mazzuchello L.L. 1075
McCoy A.B. 1004
McGee-Lennon M. 487
McGowan C.C. 994
McGregor C. 453
Medeiros Borges M. 1007
Medeiros G.F. 800
Mehrabi S. 604, 1033
Mei J. 300, 711
Mejía-Zapata D. 1062
Melo M.T.D. 1104
Melton G.B. 158, 706, 785, 821
Meltzer D. 584
Mendonça Queiroga R. 938
Meneghetti J.C. 653
Menezes Jr. J. 959, 970
Merabti T. 544, 1036
Merzweiler A. 222
Messina L.A. 1010
Metcalfe P.D. 1065
Metzger M.-H. 1067
Meystre S.M. 599, 609
Miñarro-Giménez J.A. 716
Michalowski W. 1100
Mihalas G.I. 1041, 1044
Milano F.E. 672, 1025, 1026
Millar J. 790
Mills C. 320
Min L. 207
Minelli Figueira R. 991
Mitsuhiro T. 1120
Miyabe M. 511, 1113
Miyahira A. 944
Miyamoto T. 944
Mo H. 1098
Mohyuddin 1053, 1054
Monje-Ortega D. 329
Monkman H. 338, 358
Monteiro Grendene G. 990
Monthaler P. 227
Monti M. 280
Morais M. 227
Morales F. 925
Morassi Sasso A. 534
Moreira Lucena A. 986
Moreno C. 942
Morgan S. 7
Moriguchi S. 911
Morino K. 902
Moriyama M. 79
Moro C. 936
Moro C.M.C. 952
Moser S.E. 614
Motiwala T. 414
Mougin F. 815
Moura J.A.B. 857
Muchaluat-Saadea D.C. 746
Mukai M. 911, 933
Mukherjee S. 1055
Mulwa E. 780
Münch H. 893
Munro C. 320
Murphree D. 721

Murrali M. 629
Murray P. 501
Murta A.S. 800
Muscolo D.L. 672
Muturi D. 677
Muxí C. 921
Myneni S. 113

N
Nagaie S. 1057
Nagao H. 944
Nakagawa Y. 944
Nakamura N. 1080
Nakashima N. 79, 975, 1073, 1077
Nakashima Y. 1077
Nakaya J. 1057, 1080
Nakayama M. 902, 1080
Naumann R. 909
Navas H. 903
Navathe A.S. 629
Navathe S. 629
Nemeth H. 12
Neriz L. 960
Neto J.A. 227
Neves D.S. 989
Neylan T. 736
Ng K. 1056
Ngufor C. 721
Nguni R. 677
Nguyen L. 912
Ni J. 35
Niang M. 1018
Nicholson W. 946
Nickel S. 1066
Nicol E. 993, 1000
Nienhoff H. 1072
Nikolski M. 1005
Nin M. 372
Nishikitani M. 975
Niyibizi G. 880
Nøhr C. 497
Noboa O. 372
Nogueira Reis Z.S. 935
Nohara Y. 79, 975, 1073
Noma H. 198
Norén G.N. 574
Noronha I.L. 372
Norris T. 1008
Northern California HIE Working Group 931
Novaes M.A. 1108
Nshimyiryo A. 880
Nunes Ferreira R. 937
Núñez Mondaca A. 1105
Nüssli S. 1011
Nyssen M. 193, 482, 525
Nytrø Ø. 954, 980

O
Oberpaur B. 1001
Occhiuzzi R. 921
Ochoa S. 396
Ofoghi B. 1102
Ogallo W. 1088, 1106
Ogishima S. 1057
Ohmann C. 1005
Ojwang J. 677
Okamoto K. 926
Oksen D.V. 1005
Okuda Y. 911, 933
Oleynik M. 534, 1040
Oliveira Rodrigues É. 726
Oliveira A.E.F. 847
Oliveira I. 976
Oliveira I.J.R. 935
Oliveira J.L. 687
Oliveira L. 1022, 1027, 1028
Oliveira M.A. 653
Oluoch T. 677
Ong M.-S. 761
Onofre de Lira C. 940
Opelz G. 1049
Orellana García A. 310
Ortiz de Camargo M. 940
Ortiz D.C.F. 943
Ortiz D.R. 943
Ortiz J.M. 368
Osterhage K. 978
Otero C. 368, 434, 877, 967
Otero P. 434
Othman S.B. 305
Ouyang P. 89
Overhage J.M. 60
Oza S. 916

P
Pachéco J. 212
Pacheco de Oliveira S. 995
Paciello J. 396
Padman R. 315, 1063, 1081
Paese F. 776, 941
Pageler N. 918
Pahontu R. 222
Pakhomov S.V.S. 785, 821
Palakal M. 604
Palhares D.M.F. 69
Palma J. 918, 931
Palomba E.B. 648
Pamponet Machado A. 800, 1020
Panahiazar M. 40
Pape J. 883
Parada D. 960
Paralescu S. 1041, 1044
Paranjape R. 1081
Parimbelli E. 998
Park H.-A. 142, 1099
Park R.W. 574
Parkulo M. 946
Parry D. 968, 1008
Pathak J. 40, 721, 946, 1093, 1097, 1098
Patrão D.F.C. 534, 1040
Patterson O.V. 614, 1035
Payne P.R.O. 414, 663, 1037, 1071, 1107
Pazos Gutiérrez P. 45
Pazos-Arias J. 50
Peabody A.L. 1037
Pearce C. 458
Pedersen Å.-M. 878
Pederson R. 934
Peek N. 320, 419, 424, 696, 701
Pefaur J. 372
Pelayo S. 353
Pereira Afonso dos Santos J. 987
Pereira N. 40
Pereira S. 275
Peres Penteado A. 1058
Peres H.H.C. 943
Peres L.M. 827
Pérez Alfonso D. 310
Périgard C. 212
Perkusich A. 549
Perkusich M. 549, 862
Perodin C. 883
Perreault J.O. 842
Persoz C. 1005
Pessoa C.G. 988, 989
Pettersen Nytun J. 891
Peura S. 982
Phua J. 584
Pieczkiewicz D.S. 401
Pillai P.S. 731
Pina C. 970
Pinheiro da Rosa R. 1045
Pinheiro H.M. 906
Pinto Pizzo V.R. 940
Pistarini C. 998
Platchek T. 918
Plazzotta F. 386, 890
Plischke M. 909, 1119
Plössnig M. 950
Plotnicki L. 1049
Plouin P.-F. 217
Podchiyska T. 1084
Polubriaginof F. 907
Poncelet P. 137
Poole S. 918
Porn A.M. 827
Portilla F. 434
Pratt N. 574
Preston S. 123
Prey J.E. 907
Price M. 237, 928
Price R. 584
Prokosch H.-U. 1005
Prosperi M. 701

Przysucha M. 1072
Psaraki M. 1117
Puła A. 325
Pyrz K. 320

Q
Qian Q. 1061
Qian Y. 1027, 1028
Quade M. 492
Quaglini S. 998
Quilliot A. 305
Quintana Y. 977
Quisi-Peralta D. 50
Quispe M. 967

R
Rafael N. 1052
Rahmel A. 1049
Raje S. 663, 1107
Ramanathan J. 1107
Ramaprasad A. 1105
Ramirez-Robles M. 1018
Ramis F. 960
Ramos Enck C. 1007
Ramos G. 916
Ran Kim A. 1099
Rance B. 771
Randell R. 443
Raniery Ferreira Junior J. 1079
Rao P. 946
Rappelsberger A. 295, 950, 1111
Rasmussen L.V. 1098
Rastegar-Mojarad M. 539, 1033
Ravvaz K. 1124
Ray P. 1069
Rea S. 21
Rebeiro Hargrave A. 79
Recondo F.J. 255
Rector A. 790
Redda D. 639
Redley B. 912
Reeza Mustaffa H. 1090
Regan K. 414, 663
Reich C.G. 574
Reichert F. 891
Reis S. 506
Reis Z.S.N. 936, 976
Remera E. 880
Renard J.-M. 305
Renault E. 913
Restrepo N. 1062
Reytarowski J. 909
Rhienmora P. 163
Ribeiro A.L. 69, 991
Ribeiro-Alves M.A. 372
Ricarte I.L.M. 55
Richards J. 677
Richter C. 909
Rienhoff O. 492
Righi L.V. 1121
Rijnbeek P.R. 574
Rindflesch T.C. 1094
Ringler M.D. 922
Risk M. 967
Ritacco L.E. 672, 1025, 1026
Rizzato Lede D.A. 232
Roberts C. 123
Robinson D. 790
Robles-Bykbaev V. 50, 329
Roch A.M. 604
Rocha R. 7, 805
Rocha R.A. 629
Rochon M. 909
Rodas E.B. 329
Rodrigues J.-M. 790, 1036, 1090
Roesler V. 648
Roger A. 672
Rogers M. 406
Rognoni C. 998
Rohrer Vitek C. 946
Roitman P.D. 672
Rosenman M.B. 60, 1009
Ross J. 736
Rosu D. 1100
Rothenberger D.A. 821
Rottmann T. 492
Roudsari A. 1110
Rousseau A. 264
Roussi K. 1059
Roy P.C. 118
Ruau D. 1065
Ruddle R.A. 443
Ruibal Faral K. 923
Russell A. 148
Ryan P.B. 574

S
Saba V. 776
Sacaleanu B. 1032
Sacchi L. 682
Sacoto H. 329
Sadhu E. 584
Safran C. 501, 977
Saini S.D. 614
Salamea J.C. 329
Salkeld G. 905
Salokannel M. 1005
Salvini R. 741
Samuel A. 964
Samuel U. 1049
Samuelson K. 736
Sandefer R.H. 401
Sanders C. 123
Sangare M. 1018
Santana F. 1108
Santojanni A.M. 472
Santos Alves D. 64
Santos Neves D. 987
Santos Simões de Almeida L.H. 929
Santos A.A. 857
Santos F.A.O. 589
Santos M.L.O. 746
Saraiva R.M. 862
Saranto K. v, 1014, 1116
Saravanamuthu C. 663
Sarkar I.N. vii
Sarti T.D. 888
Sasaki S. 198
Savadogo M. 1103
Savard S. 217
Sax U. 668
Sayvong R. 429
Scandurra I. 153, 908
Schaarup C. 84
Schachner M.B. 255
Schaffer J.L. 183
Schedel F. 1042
Scheppmann D. 158
Schimidt M.Q. 888
Schleder Gonçalves L. 1012
Schlegel D.R. 619
Schmidt H. 604
Schmidt M. 604
Schnall R. 467
Schneider G. 222
Schreiweis B. 884
Schrom H. 1060
Schuemie M.J. 574
Schuers M. 212
Schuh C. 1119
Schuler J. 950
Schulte G. 492
Schultz C. 946
Schulz S. 716, 790, 832, 909, 1108
Schulze Sünninghausen S. 837
Schwab P.K. 852
Schwartze J. 909, 1060
Schweitzer M. 1118
Seeling W. 1119
Segagni D. 682
Seger D.L. 242
Segond F. 1067
Seki M. 933
Sellemann B. 492
Semic-Jusufagic A. 320
Semple J. 894
Senturk I.F. 766
Sequeiros G.O. 746
Séroussi B. 264, 275
Serrano J.A. 74
Seto R. 961
Sexton J. 448
Shabtai I. 1063

Shachak A. 506
Shah N.H. 574
Shahar Y. 594
Shaoa Y. 639
Sharek P. 918
Sharp C. 931
Sheets L. 969
Shikata S. 511, 1113
Shimamoto Y. 511
Shimizu S. 79
Shimokawa K. 1057
Shin S.-Y. 882, 924
Shin Y. 924
Shiratori N. 1080
Shorten R. 462
Shrestha P. 163
Shunsuke D. 1120
Shusaku T. 1120
Siebra C. 862
Siggaard Mathiesen K. 1015
Sigulem D. 168
Siika A. 520
Silva Layes E. 925
Silva E.A. 800
Silva L.C. 549
Silva L.F. 746
Silvariño R. 372
Silverstein J.C. 584
Simón M. 368
Sim I. 736
Simon G.J. 706
Simpson A. 320
Sims J. 653
Singh K. 901
Sinha M. 953
Sittig D.F. 173, 1004
Skyrme A. 506
Slaughter L. 954
Smaradottir B. 892
Smith D.E. 1026
Smith J. 1062
Smith M.I. 232
Soares Garcia V. 986
Soares de Vasconcelos G. 1125
Sockolow P.S. 406
Sohn S. 1033
Solbrig H. 790
Solbrig H.R. 1097, 1098
Solomonides A. 584
Some B.M.J. 1103
Sommer J. 890
Sommer J.A. 255
Song T.-M. 1099
Sordo M. 629
Soto M. 875, 885
Soto-Rey I. 1029
Sousa L. 988
Soussa V. 1059
Souvignet J. 1090
Souza M.A. 1047
Souza T.A. 1047
Spackman K. 790
Spalding S.J. 554
Spano J.-P. 264
Speedie S.M. 401
Speltz P. 1098
Spencer S. 1065
Sperrin M. 696, 701
Ständer S. 1042
Staemmler M. 893
Stang P.E. 574
Stanziola E. 368
Stawiski K. 325
Steinbrügge C. 909
Stockton C. 123
Stoehr M. 259
Strauss G. 259
Strumia G. 903
Strzałka A. 325
Stub Petersen L. 376, 1002
Sturm J. 893
Suchard M.A. 574
Suebnukarn S. 163, 756
Suero-Tejeda N. 984
Suganuma T. 1080
Sugimoto M. 79
Sun J. 1051
Sun L. 974, 1024
Sun Y. 1109
Sung M.-W. 904
Suomi R. 982
Süsal C. 1049
Sutton J. 946
Suzuki K. 1050
Sylvestre E. 913
Syn T. 1105
Szczecinski Rodrigues Á. 990

T
Tada K. 944
Takahashi S. 751
Takahiro S. 1120
Takai T. 1057
Takaoka S. 1068
Takase K. 198
Takeda H. 1017
Takemura T. 198, 926, 944, 966
Takeo Ueda E. 937
Takeuchi H. 751
Talavera J. 396
Tam E. 1086
Tamayo-Correa C. 1062
Tambyraja R. 158
Tanús R. 372
Tan Z.-H. 84
Tanaka H. 1057
Tao C. 113, 1051, 1076
Tarlov E. 584
Tarrier N. 123
Taslimitehrani V. 40
Teich J.M. 916
Teixeira Macedo H. 589
Tellez M. 247, 943
Tellis R. 1027, 1028
Tenório J.M. 1089
Tenschert J. 852
Thanathornwong B. 756
Themis M. 320
Thiebaut R. 1005
Thompson C. 886, 1056
Thompson H.J. 964
Thompson W.K. 1098
Thorne N. 1052
Thornton S.N. 21
Thyvalikakath T.P. 1081
Tiase V. 907
Tibollo V. 682
Tierney D. 7
Tilahun B. 927, 981
Timbi-Sisalima C. 329
Ting W.H. 945
Tizatto L.A.P. 648
Tominaga T. 1080
Tong L.H. 945
Toniazzo de Abreu L.L. 1074, 1075
Tönshoff B. 1049
Topaz M. 242, 629
Torkilsheyggi A. á 477
Torres Pisa I. 168, 1010, 1058, 1089
Torres-Silva E.A. 1062
Torresani M. 280
Toubiana L. 217
Tournigand C. 264
Toussaint P. 103
Touzé E. 217
Traore S.T. 1018
Treanor D. 443
Trick W. 584
Tripp J.S. 270
Trombert B. 1090
Trovato K. 1027, 1028
Truong T. 1101
Tsafnat G. 761
Tsiknakis M. 1117
Tsuji H. 933
Tu W. 60
Tucker S. 999
Tudor A. 1041, 1044
Turner A.M. 380, 978, 979
Turner M. 1052
Tvardik N. 1067

U
Ueda T. 1068
Ugon A. 217
Umapathy H. 905

Umbach N. 668
Upadhyaya S. 721
Üstün B. 790
Usuda Y. 1113
Uznayo M.Q. 368

V
Van Den Bergh R. 920
van der Lei J. 574, 687
van der Veer S.N. 424
van Engen-Verheul M.M. 424
van Staa T. 1005
Van Woensel W. 118, 148
Vandenbussche P.-Y. 264
Varghese J. 837
Vawdrey D.K. 907
Vecellio E. 917
Veiga Silva A.C. 1075
Veil K. 458
Vélez-Zuluga S. 1062
Venson R. 1075, 1082
Verbeke F. 193, 482, 525
Verger P. 1005
Vernon M. 448
Vieira Esteves T. 995
Vieira V.J.D. 1047
Viernes B. 1035
Villalba C. 132
Villumsen S. 497
Virginio Jr. L.A. 55
Vitti S.V. 168
Vivent M. 1001
von Cavallar S. 691
von Heideken J. 901
von Ingelheim J. 909
von Kaenel F. 1011
Vyawahare K. 1081

W
Wagner M. 909
Wahl A.M. 852
Wai A. 1086
Wakabayashi S. 961
Wakata Y. 1073, 1077
Waks Z. 280
Waleska Simões P. 1074, 1075, 1082
Wang D. 515
Wang L. 207
Wang S. 569
Wang Y. 706, 821, 1024
Wanyee S. 780
Waruru A. 677
Watabe E. 1113
Watbled L. 212
Waters K.P. 677
Watson N. 487
Weaver F.M. 584
Weber J.H. 237, 928
Weiner M. 736
Weng C. 569, 1023
Were M.C. 520, 780
Westbrook J.I. 917
Wetter T. 951, 973
Whelan P. 123
White P. 1110
Wickner P.G. 242
Wickramasinghe N. 183, 912
Wilk S. 1100
Williams J. 609
Williams P. 458
Williams R. 696
Winnenburg R. 795
Wipfli R. 898
Witter Z. 766
Wix K. 946
Wolf K.-H. 909, 1060
Wong A. 1086
Wong C.M. 945
Wong I.C.K. 574
Wood J. 89
Wortley S. 905
Wright A. 1004
Wright G. 525
Wu S.Y. 945
Wu Y. 624, 899, 974, 1024
Wubin Q. 1083
Wykes T. 123

X
Xiao Q. 1024
Xiao S. 974
Xiaolei W. 1083
Xie G. 300, 711
Xu H. 624
Xu S. 899
Xu Y. (Yahong) 974, 1024
Xu Y. (Yimin) 974

Y
Yamada H. 926
Yamashita T. 1077
Yang F. 899
Yang J. 639
Yang L. 1061
Yang Q. 974
Yang X. 899
Yao Y. 1035
Yasini M. 127
Yasushi M. 1120
Yokooka Y. 933
Yoon S. 984
Yoshida Y. 911
Yoshihara H. 881, 1050
Yu Y. 300, 711, 1033
Yue P. 974
Yun Lee J. 1099
Yutaka H. 1120

Z
Zanen W. 1049
Ze Z. 123
Zekri O. 564
Zelek L. 264
Zelisko S. 584
Zeng-Treitler Q. 639
Zhang H. 584
Zhang J. 965, 1085
Zhang P. 1056
Zhang Q. 1017
Zhang R. 529, 785
Zhang Y. 1031
Zhong F. 629
Zhou L. 242, 629
Zhou X. 761, 1085
Zhu M. 965
Zhu V.J. 60
Zhu X. 1022, 1056
Zhu Y. 1085
Ziegenspeck A. 909
Ziv A. 506
Zosso A. 898

This page intentionally left blank

This page intentionally left blank