Conservation Biology

Essays

Making literature reviews more reliable through application of lessons from systematic reviews

N.R. Haddaway,* P. Woodcock,† B. Macura,‡§ and A. Collins**

*MISTRA EviEM, Royal Swedish Academy of Sciences, Lilla Frescativägen 4A, 114 18, Stockholm, Sweden, email neal_haddaway@hotmail.com
†JNCC, Monkstone House, Peterborough, PE1 1JY, United Kingdom
‡Land, Environment, Agriculture and Forestry Department, University of Padova, 35020, Legnaro, Padova, Italy
§Centre for Evidence-Based Conservation, School of Environment, Natural Resources and Geography, Bangor University, Bangor, LL57 2UW, United Kingdom
**Centre for Environmental Policy, Imperial College London, London, SW7 1NA, U.K.

Abstract: *Review articles can provide valuable summaries of the ever-increasing volume of primary research in conservation biology. Where findings may influence important resource-allocation decisions in policy or practice, there is a need for a high degree of reliability when reviewing evidence. However, traditional literature reviews are susceptible to a number of biases during the identification, selection, and synthesis of included studies (e.g., publication bias, selection bias, and vote counting). Systematic reviews, pioneered in medicine and translated into conservation in 2006, address these issues through a strict methodology that aims to maximize transparency, objectivity, and repeatability. Systematic reviews will always be the gold standard for reliable synthesis of evidence. However, traditional literature reviews remain popular and will continue to be valuable where systematic reviews are not feasible. Where traditional reviews are used, lessons can be taken from systematic reviews and applied to traditional reviews in order to increase their reliability. Certain key aspects of systematic review methods that can be used in a context-specific manner in traditional reviews include focusing on mitigating bias; increasing transparency, consistency, and objectivity, and critically appraising the evidence and avoiding vote counting. In situations where conducting a full systematic review is not feasible, the proposed approach to reviewing evidence in a more systematic way can substantially improve the reliability of review findings, providing a time- and resource-efficient means of maximizing the value of traditional reviews. These methods are aimed particularly at those conducting literature reviews where systematic review is not feasible, for example, for graduate students, single reviewers, or small organizations.*

Keywords: evidence assessments, evidence reviews, meta-analysis, rapid assessments, rapid reviews

Revisiones Bibliográficas más Confiables Mediante la Aplicación de Lecciones de las Revisiones Sistemáticas

Resumen: *Los artículos de revisión pueden proporcionar resúmenes valiosos del volumen siempre creciente de la investigación primaria en la biología de la conservación. Cuando los hallazgos pueden influenciar las decisiones importantes de asignación de recursos en la política o en la práctica, entonces existe una necesidad de un alto grado de confiabilidad cuando se revisa la evidencia. Sin embargo, las revisiones bibliográficas tradicionales son susceptibles a un número de sesgos durante la identificación, selección y síntesis de los estudios incluidos (p. ej.: sesgos de publicación, de selección y el conteo de votos). Las revisiones sistemáticas, aplicadas primero en la medicina y después llevadas a la conservación en 2006, se dirigen a estos sucesos por medio de una metodología estricta que busca maximizar la transparencia, la objetividad y lo repetible. Las revisiones sistemáticas siempre serán el máximo estándar para la síntesis confiable de la evidencia. Sin embargo, las revisiones bibliográficas tradicionales siguen siendo populares y seguirán siendo valiosas cuando las revisiones sistemáticas no son factibles. Cuando se usan revisiones tradicionales, se pueden tomar lecciones a partir de las revisiones sistemáticas y aplicarlas a las revisiones tradicionales para incrementar su confiabilidad. Ciertos aspectos claves de los métodos de las revisiones sistemáticas que pueden usarse de*

Paper submitted October 5, 2014; revised manuscript accepted March 11, 2015.

Conservation Biology, Volume 00, No. 0, 1–10

DOI: 10.1111/cobi.12541

manera específica para el contexto en las revisiones tradicionales incluyen enfocarse en el sesgo de mitigación, incrementar la transparencia, la consistencia y la objetividad, y valorar la evidencia y evitar el conteo de votos. En las situaciones en las que no es factible realizar una revisión sistemática plena, la estrategia propuesta de revisar la evidencia de manera más sistemática puede mejorar sustancialmente la confiabilidad de los hallazgos de las revisiones, lo que proporciona medios eficientes en tiempo y recursos para maximizar el valor de las revisiones tradicionales. Estos métodos están dirigidos particularmente a aquellos que realizan revisiones bibliográficas en los que una revisión sistemática no es factible, por ejemplo, los estudiantes de posgrado, un revisor único, u organizaciones pequeñas.

Palabras Clave: evaluaciones de evidencia, evaluaciones rápidas, meta-análisis, revisiones de evidencia, revisiones rápidas

Introduction

Research is being published at an ever-increasing rate (Larsen & von Ins 2010; Pautasso 2012); both the number of new publications and new journals is increasing rapidly (Michels & Schmoch 2012). Review articles that synthesize primary research can therefore be extremely valuable because they summarize this expanding body of evidence, clarify controversies, and identify research gaps. Importantly, readers of reviews must trust that the reviewers' conclusions are reliable and that the literature is summarized in an unbiased way. However, unintentional bias can creep into reviews and lead to inaccuracies or misinterpretation of the evidence (Gates 2002; Roberts et al. 2006; Philibert et al. 2012). Indeed, traditional literature reviews are often based on selected and possibly limited sources, focusing on studies with which the reviewers are already familiar (Petticrew & Roberts 2008).

Because review findings are increasingly used to inform decisions with important environmental and socioeconomic implications (e.g., Halme et al. 2010), they need to be particularly reliable. The need for rigor and clear accountability in decision making has resulted in the development and adoption of systematic reviews. Systematic reviews differ from traditional literature reviews in that a series of strict guidelines must be followed to ensure that the influence of subjective judgment is minimized, that the review process is fully repeatable, and that the review is as comprehensive as possible. For example, systematic reviews include extensive searches for unpublished literature to overcome possible publication bias (the tendency for non-significant or controversial results to remain unpublished), and multiple academic databases are searched to minimize the likelihood of relevant studies being missed (Petticrew 2001; CEE 2013). Systematic review guidance for conservation was adapted from similar guidance in medicine in 2006 (Pullin & Stewart 2006) and was shortly followed by the establishment of the Collaboration for Environmental Evidence (CEE), an organization that governs the publication of and guidance for systematic reviews in environmental science. Over the past 15 years, the number of systematic reviews and studies using associated techniques, such as meta-analyses, in conservation has been increasing (Fig. 1). Over 60 CEE systematic reviews have been published, and a large number are underway (see www.environmentalevidence.org). The level of objectivity and rigor in systematic review methods is increasingly regarded as the gold standard in evidence review; several decision-making organizations have commissioned systematic reviews to answer contentious questions of high policy importance (e.g., Randall & James 2012; Pullin et al. 2013; Haddaway et al. 2014*a*).

Despite the many advantages of systematic reviews, the approach is time- and resource-intensive because of the requirement to coordinate multiple reviewers, process large numbers of search results, and involve a team of expert advisors. Indeed, the duration of a typical systematic review is 9 to 24 months, with costs ranging from £20,000 to £200,000 (McGowan & Sampson 2005; CEE 2013). The quality and reliability of the resultant review can represent a sound investment. However, such resource demands can make following full systematic review guidelines prohibitive for researchers or organizations operating on limited budgets, tight time frames, or in situations where the review topic falls outside the current priorities of potential commissioning bodies.

When Traditional Reviews Are Valuable

Considerable effort is being made to expedite systematic reviews whilst maintaining reliability and rigor, and processes that establish succinct, narrow, and focused review topics are important for ensuring that systematic reviews take less time (Haddaway et al. 2015). However, traditional reviews will continue to be valuable in at least 6 situations: when a systematic review team cannot be formed, when resources are limited, when time is severely limited, where systematic review is not an accepted method, where aspects of a systematic review are unnecessary, and when the review topic does not warrant a full systematic review. When a reviewer is working alone or in a restricted group, establishment of the required review team or steering group is not feasible, and organizations with limited resources (e.g.,

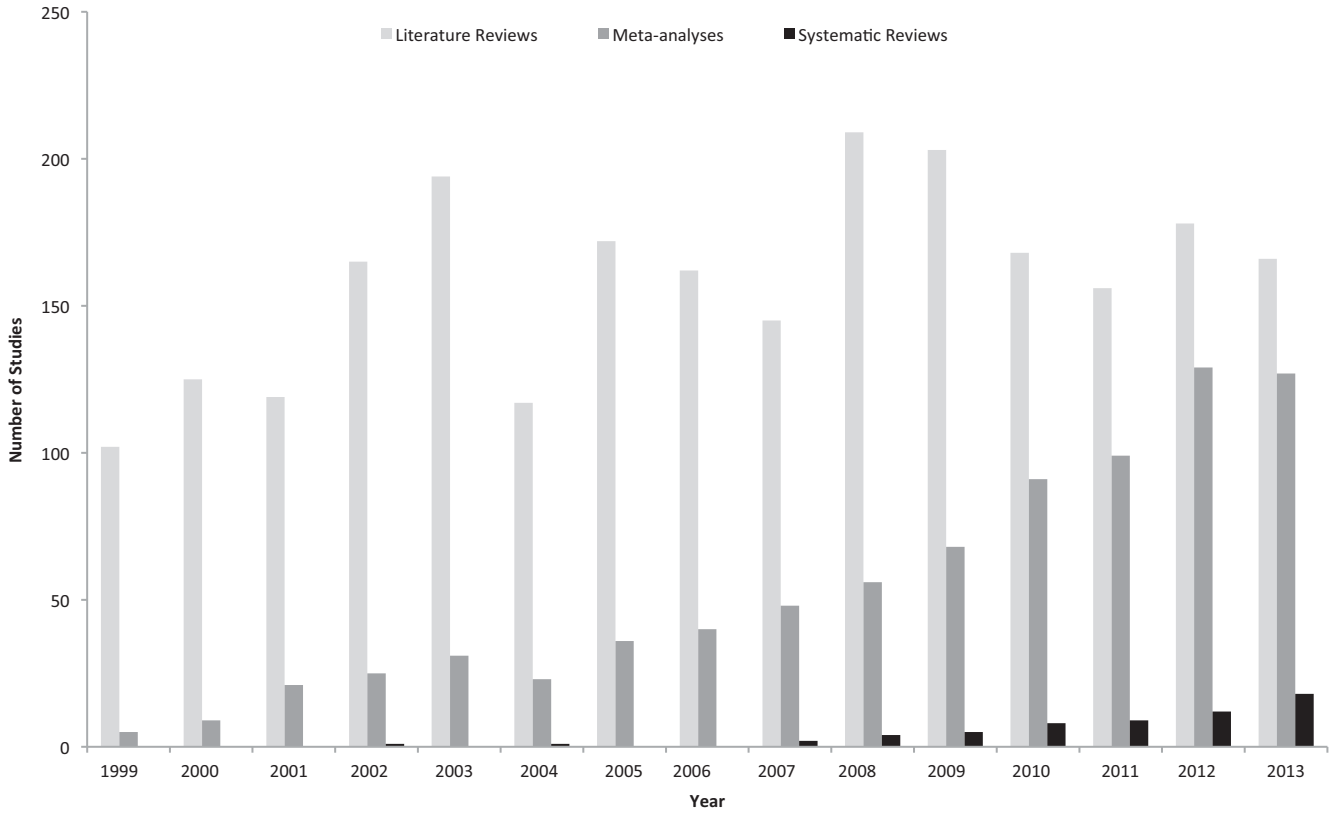

Figure 1. Number of search records based on topic word searches from Web of Science (including Biological Abstracts, MEDLINE, SciELO, Web of Science Core Collections, and Zoological Record) for literature reviews ("literature review" OR "review of the literature" OR "review of literature"), meta-analyses ("meta-analysis" OR metaanalysis OR "meta analysis"), and systematic reviews ("systematic review" [not including taxonomic systematic review or methodology/discussion papers, which were removed by hand]) in the subject category biodiversity conservation separated by year.

those with annual budgets comparable to the costs of a systematic review) cannot afford to conduct systematic reviews. Such researchers and organizations represent an important part of the global conservation network and must balance reviewing evidence with competing priorities (e.g., primary data collection, management interventions, communication and outreach) under considerable resource constraints.

Traditional reviews are needed when review commissioners are under extremely strict time requirements and a full systematic review with the associated delays during peer-review and publication are a greater concern than the reduced reliability of a traditional review, and in situations where systematic review is not an accepted method but reviewers wish to increase the reliability of their reviews as much as possible. A traditional review is also in order when some aspects of systematic review methods are unnecessary for the desired outputs of the review. For example, a traditional review is appropriate if a review need not be exhaustive, as in exploratory and configurative reviews (i.e., hypothesis- and conceptual-framework- forming reviews). These types of reviews typically reach information saturation, and the additional material from a systematic review would not add to the understanding of the topic (Gough et al. 2012). These topics may not be suitable for systematic review, but they may greatly benefit from selected aspects of the systematic review methods, such as systematic searching and screening and critical appraisal. Traditional reviews are also warranted when a review topic is too basic for full systematic review (e.g., existence of species X in region Y), although even this type of review may benefit from some systematic approaches.

Systematic Review Lessons

Based on our experience of conducting systematic reviews and using related methods, such as Civil Service Rapid Evidence Assessments and Quick Scoping Reviews (Civil Service 2010; Collins et al. 2014), we aimed to highlight key aspects of systematic review methods that can be adapted and applied to traditional reviews. Our objective was not to rewrite or reiterate

systematic review guidelines or to provide a shortcut to the systematic review process, and we emphasize to policy makers that the resulting traditional reviews will not reach the same level of rigor as a full systematic review. Rather, we identified approaches to reviewing evidence that will make important improvements to the reliability of review findings in situations where a full systematic review is not appropriate or not feasible, and we hope this guidance will help ensure that the value of traditional reviews is maximized.

Detailed descriptions and rationales for systematic review guidance are provided by the major coordinating bodies (e.g., The Cochrane Collaboration 2011; CEE 2013). These guidelines emphasize that systematic reviews should include the following key stages and components. First, prior to the commencement of the review, the method (i.e., protocol) to be used in the review is peer reviewed and published. This protocol outlines the questions to be addressed and the approaches that will be followed in the review. Second, multiple databases and sources of gray literature are searched for relevant evidence with search strings defined in the protocol. Third, all search results are screened using predefined inclusion criteria to determine whether or not articles are relevant to include in the review. Article relevance is typically assessed at progressively greater levels of detail (e.g., titles, then abstracts, and finally the full text). Fourth, screening is undertaken by multiple reviewers to ensure inclusion criteria are being followed consistently. Fifth, the methodological rigor of each primary research article is critically appraised in a transparent and objective manner. Sixth, evidence is synthesized and described, typically including implications for policy, practice, and research. Seventh, supplementary information is provided that transparently documents review activities to ensure these activities are repeatable and verifiable.

Systematic reviews typically focus on synthesizing evidence of the effect of X on Y (e.g., the effect of ungulate grazing on invertebrate diversity) or on assessing what evidence exists on a specific topic, also referred to as systematic maps (e.g., what evidence exists on the impacts of global warming on Himalayan glacial melt). Traditional reviews framed around such questions are therefore likely to derive the greatest benefit from applying systematic review methods. However, configurative reviews that aim to provide an overview of the literature in a way that minimizes potential bias can also benefit considerably from incorporating elements of systematic review guidance.

We considered the key potential limitations of traditional reviews, provide a brief description of the approach to mitigating these problems as recommended by systematic review guidance, and suggest how this guidance could be adapted to increase reliability in traditional reviews. Specifically, we focused on mitigating bias; increasing transparency, consistency, and procedural objectivity; and critically appraising the evidence and avoiding vote counting.

Estimates of the likely resource requirements and applicability of approaches to improving traditional reviews are provided in Table 1. Where a systematic review is not feasible or appropriate, we recommend that these measures be selected based on available resources and relevance to the review topic.

Mitigating Bias

Bias in reviews can involve the use of a non-representative portion of the literature or preconceptions and opinions that influence data synthesis. Biases are often unintentional, but they represent a potentially important problem for traditional reviews, particularly where contentious questions with conflicting evidence are addressed. The most prevalent forms of bias in reviews are typically selection bias and publication bias.

Selection bias arises from the inclusion of a biased selection and discussion of evidence (Table 1 points 2, 3, 4, 6, and 9), most commonly through the purposeful selection of articles to synthesize. Selection bias may also result from the selection and searching of databases that cover a non-representative subset of the literature, or from the ad hoc inclusion of studies based on reviewer awareness and familiarity (Egger & Smith 1998). Additionally, searching databases using an inappropriate search string may mean vital synonyms and potentially influential, relevant research are missed. Systematic review guidance (e.g., CEE 2013) recommends that searches for evidence use a predefined and tested search string; that all search results be screened for relevance based on predetermined criteria in order to avoid purposeful selection of a potentially biased group of studies; and that multiple databases be used to avoid potential biases that arise from individual databases cataloging a non-representative portion of the literature. Systematic review approaches thus also increase comprehensiveness.

We suggest that traditional reviews apply carefully designed search strings with appropriate synonyms and combinations of search terms and that the relevance of all search results be determined based on consistent criteria. Where resources allow, searches should also be conducted in multiple databases with a range of coverage in terms of publication dates and subject. In practice, however, searching the 5–10 databases (or more) often seen in systematic reviews (e.g., Bernes et al. 2013*a*, 2013*b*; Macura et al. 2013) might provide limited additional benefits over a more restricted search of 2–3 databases, particularly given the time requirements, the need to obtain access to relevant databases, and overlaps and redundancy among database content.

Table 1. Potential limitations of traditional literature reviews and suggested solutions.

Limitation	*Consequence*	*Frequency of limitation*[a]	*Suggested solution*	*Level of resources required for solution*[b]	*Applicability of solution*
1. Question not clearly defined	Review boundaries are subjective and susceptible to post hoc changes. Review has limited value for readers interested in specific questions.	moderate	Plan questions, desired outputs, and inclusion criteria before starting, and consult with subject experts.	low	all reviews (especially where reviewer is not an expert)
2. Database chosen for searching is unrepresentative	Susceptibility to bias: review is not comprehensive and synthesized studies are not representative.	low	Include all viewpoints by searching multiple databases.	high	all reviews
3. Search strategy and approach to deciding whether or not articles are relevant for inclusion in review is unclear	Low transparency: the literature search and screening cannot be evaluated or repeated, limiting confidence in review conclusions.	moderate-high	Include all details of searches conducted in supplementary information. Consider checking agreement over inclusion of articles with a colleague or expert for a small subset of search results.	low	all reviews
4. Search strategy misses key relevant articles	Review fails to include influential research and is thus void.	low	Spend time developing an appropriate, comprehensive search string (defining PICO [population, intervention, comparator, outcome] elements where appropriate). Test comprehensiveness of searches based on a list of known articles provided by subject experts.	moderate	all reviews (especially where reviewer is not an expert)
5. Published evidence is biased	Susceptibility to bias: review tends toward positive or significant results.	high	Include searches for unpublished (gray) literature to account for publication bias.	moderate	most reviews

Continued

Table 1. Continued

Limitation	*Consequence*	*Frequency of limitation*[a]	*Suggested solution*	*Level of resources required for solution*[b]	*Applicability of solution*
6. Selection of articles is subjective or purposeful	Susceptibility to bias: synthesized studies are not representative.	moderate-high	Screen all search results with the same predetermined criteria.	high	all reviews
7. Failure to consider variation in study quality	Poor studies are given as much weight as good studies.	moderate-high	Create basic critical appraisal tools to categorize articles as of low, high, or unclear quality. In quantitative analysis (meta-analysis), weight studies and perform sensitivity analyses to investigate influence of low quality studies.	high	all reviews (meta-analysis where feasible)
8. Failure to consider magnitude of study effects	Significance or magnitude of effect sizes for individual studies is not provided.	moderate	Apply weighted, quantitative syntheses (meta-analysis) where possible. For qualitative analyses describe direction and magnitude of significance, including detectability (sample size or background noise vs. signal).	low	all reviews (meta-analysis where feasible)
9. Review result or discussion focuses mainly on specific studies rather than describing the evidence base as a whole	Results may be biased by subjective weighting of certain studies that do not warrant more attention than others.	high	Attempt to give equal weighting to all studies of similar quality and relevance. Clearly state why some studies are not discussed at length (e.g., not relevant or low quality). Describe evidence base as a whole where possible (i.e., in figures, tables, and meta-analyses).	low	all reviews

[a]*Estimate based on evaluations of the rigor of traditional reviews in ecology and conservation (Roberts et al. 2006; Huntington 2011; Koricheva & Gurevitch 2014; Woodcock et al. 2014).*
[b]*Estimated based on the authors' experience in conducting systematic reviews.*

Publication bias reflects the propensity for journals to publish studies with positive, hypothesis-affirming, or significant results rather than negative, contentious, or non-significant findings (Dickersin 1990; Lortie et al. 2007). Accordingly, including only peer-reviewed publications in a review can lead to overestimation of an effect (Table 1, point 5), in some instances with important implications for review findings (Møller & Jennions 2001; Sterne et al. 2001). Although it is difficult to assess the presence of publication bias definitively (Torgerson 2006), systematic review guidance recommends reducing the risk of publication bias through comprehensive searches for gray literature (e.g., unpublished theses, organization reports, government papers). However, gray literature searches in systematic reviews typically involve consulting tens of websites of relevant organizations as well as internet search engines. This process can be highly time-consuming but yield variable benefits because different research fields contain different volumes of gray literature. For example, little gray literature was found in a systematic review of greenhouse gas emissions from lowland peat landscapes (Haddaway et al. 2014*a*) because measurement equipment is complex and expensive. By contrast, a systematic review of impacts of terrestrial protected areas on human well-being yielded a large quantity of evidence from gray literature (Pullin et al. 2013). For non-systematic reviews, we therefore suggest that the extent of gray literature searches be determined on a case-by-case basis, where possible in consultation with experts with experience in searching the gray literature on the topic of interest.

Reviewers should consider carefully whether relevant gray literature that meets all necessary criteria for inclusion in the review could exist, and, if so, whether this literature would be readily accessible. Reviewers should also consider what types of gray literature are likely to be important. Gray literature is of two major types: as yet unpublished academic studies and practitioner reports (i.e., studies that were never intended to be published in academic journals). These two forms of gray literature are found in different locations and efforts to obtain them should reflect these differences to ensure efficiency. For example, unpublished academic manuscripts are found in repositories of theses, conference proceedings, and pre-print servers and can be retrieved through a call for evidence aimed at the research community. Practitioner-held data can be obtained from organizational and governmental websites and libraries and through direct contact with practitioners such as environmental consultants. In addition, a reviewer could search only those organizations with websites that list or provide search facilities for publications, as opposed to news and blog articles, which in our experience rarely provide useful gray literature.

Increasing Procedural Objectivity, Consistency, and Transparency

Traditional reviews may concisely summarize evidence on a specific topic but often fail to provide details on which sources were searched for information, how studies were selected, why certain studies were excluded, or when review work was carried out (Table 1, points 1 and 3) (Roberts et al. 2006; Woodcock et al. 2014). Where the body of evidence is small and easily found, this may have limited impacts on reliability. Generally, however, lack of methodological documentation and transparency may conceal subjective choices (which may include biases, both conscious and unconscious) and means reviews cannot be repeated, evaluated, or updated. This reduces the confidence that readers place in the review conclusions and limits the future utility of the review (Koricheva & Gurevitch 2014).

To be sure, systematic reviews also involve subjective judgments. For instance, reviewers may subjectively decide on the geographic or temporal bounds of a search strategy, on criteria for determining study inclusion and exclusion, and on what thresholds (if any) should apply when extracting and synthesizing evidence (Table 1, points 3, 7, and 8). In systematic reviews, however, protocols are formulated beforehand to limit the "play of subjectivity" (Megill 2007:123), promote consistent research practices, and ensure the repeatability and reliability of important processes. For instance, coding manuals may be developed to support consistent, repeatable data extraction, and double-coding may be used to ensure agreement among reviewers. At the data synthesis stage, meta-analytic techniques can provide a rigorous framework for statistically synthesizing the results from individual studies. Comprehensive guidance is available on the application of meta-analyses in ecology (Koricheva et al. 2013), though we acknowledge that meta-analysis is not always suitable for the data available.

Because subjectivity is inevitable, transparency is paramount. Subjective choices are made explicit to a potential user by clearly stating and explaining the rationale for them. Systematic reviews also maintain transparency by publishing extensive supplementary information detailing the strategy for searching for articles, how search results were screened for relevance, why particular studies were excluded from the review, how screening decisions were checked for consistency, and how data were extracted from studies and synthesized (e.g., Haddaway et al. 2014*a*). In many cases, maintaining transparency does not increase time or costs substantially if planned from the outset because it involves reporting activities that have been undertaken already.

Critically Appraising Primary Research and Avoidance of Vote Counting

Primary research frequently varies in scope and in methodological rigor, even when similar questions are addressed. Whilst each piece of research may be sufficiently relevant for inclusion in reviews, failing to take into account variation in methods and giving equal weight to evidence of varying quality can result in unreliable or misleading conclusions (Englund et al. 1999; Gates 2002) (Table 1, points 7 and 8). Alternatively, reviews may focus on some studies rather than others, ignoring or subjectively evaluating the reliability of the individual studies (Table 1, point 9).

The problem of variable primary research quality can affect the reliability of both qualitative and quantitative traditional reviews. We define the former as reviews that apply exclusively narrative approaches to summarizing studies, whilst the latter refers to reviews that use statistical methods to combine and summarize study results. In traditional reviews (qualitative or quantitative), consideration of the methods used by individual studies is often ad hoc and subjective or absent altogether (Woodcock et al. 2014). Quantitative syntheses such as meta-analyses typically employ a more transparent approach to accounting for variation in study quality, in which each study is weighted based on a specific metric (Koricheva & Gurevitch 2014). For example, more weight is often given to studies with low variability or high sample sizes. However, whilst weighting by variance favors more precise studies, these are not necessarily more accurate. Furthermore, neither method of weighting incorporates important aspects of study design such as pseudoreplication, purposive sample selection, and inappropriate controls.

One form of vote counting occurs when all evidence is treated as equally influential in a synthesis, without regard for the reliability or variability of study findings. Another form of vote-counting occurs where synthesis is based solely on statistical significance of the included studies, for example through tallying the number of studies with significant versus non-significant findings. By failing to consider and synthesize the magnitude of effects in each individual study (the effect size), important trends across the evidence base can remain undetected and sources of heterogeneity between seemingly contradictory studies are not accounted for (Harrison 2011). Vote counting in traditional reviews can be easily avoided. Reviews should aim to extract and present study results (i.e., effect sizes, sample sizes, and variability [see Haddaway 2015]) in full rather than synthesizing studies based on statistical significance alone. Effect sizes combined with a measure of variability summarize study findings well, and their calculation is covered extensively in dedicated publications on quantitative synthesis (e.g., Borenstein et al. 2009). Reviews that do not undertake meta-analysis should still present effect sizes and variance wherever possible and avoid simple tallying of significance. Additionally, an assessment of susceptibility to bias for each included study allows reviewers to assess which evidence is most reliable.

Systematic review guidance has a strong focus on critical appraisal of primary research (CEE 2013; Bilotta et al. 2014). Critical appraisal is the process by which individual studies are assessed for external (generalizability) and internal (quality) validity (Bilotta et al. 2014). For example, methods used in several recent CEE systematic reviews (e.g., Haddaway et al. 2014*a*, 2014*b*) are based on the Cochrane Collaboration Tool for Assessing Risk of Bias (Higgins et al. 2011) and state that key factors affecting study susceptibility to bias should be considered in critical appraisal: for example, level of true replication; appropriateness of control; methodological detail; sample selection; measurement quality; presence of confounders; and study design (i.e., control and intervention, before and after, beforeafter control impacts [BACI], and randomized control trial [RCT]) . For these factors, a basic categorical scoring (e.g., 0, 1, 2) can help reviewers form a final judgment for each article (e.g., high, low, unclear quality). This judgment can then be used to inform the synthesis, for example by grouping higher quality studies together in a narrative or including study quality in meta-analysis as an explanatory variable or a basis for sensitivity analysis (see Borenstein et al. [2009] for details of meta-analysis).

For qualitative data, a variety of critical appraisal tools have been produced that can be adapted for use in conservation (Spencer et al. 2003; DfID 2014). Planning and undertaking critical appraisal using such tools can involve considerable thought and time, but the process becomes more efficient with experience and need not be a significant bottleneck if the volume of synthesizable evidence is not extensive, for example involving fewer than 100 studies. Whilst we recommend this approach, any objective, transparent assessment of quality is beneficial to all forms of review. Accordingly, where resource constraints prevent a full appraisal using these tools, evaluating at least some aspects of the methods of each study and presentation of this information can substantially improve the reliability of traditional reviews.

Further Comments

We have conducted a large number of systematic reviews and found the following to be important considerations when planning and conducting traditional reviews. In our experience, reviewers should not be scared by inexperience. Lead reviewers in systematic reviews are often not specialists in the subject area (Petticrew 2001). Although the inclusion of experts in the review team is vital, the

relative naiveté of the lead reviewer has benefits. In particular, non-specialists are less likely to be influenced by preconception bias (pre-existing opinions about which studies should be included in the review and how these studies should be interpreted) and are often better placed to appreciate the need to describe details of the review process and technical terms for other non-specialists.

Reviewers should be realistic when assessing how much can be done to make reviews more systematic and consider the potential costs and benefits of more comprehensive approaches. In this respect, we have found scoping exercises to be very useful. For example, it is possible to produce a fairly accurate estimate of the number of article titles, abstracts, and full texts that can be screened for relevance per day by examining a random sample of 100 titles. This will indicate inclusion rates at the three stages and will help one predict time needed to screen all search results. Similar scoping can also be applied to subsequent stages of the review process (e.g., critical appraisal, data synthesis) to estimate time requirements and feasibility.

Reviewers should always be aware of the potential biases in their methods (and in the primary data included), and these potential biases should be clearly noted in the review discussion.

Finally, we emphasize that systematic review methods are being widely used in science policy, research, education, public health, and medicine. These are valuable methods for reviewers, even if they do not have the resources to undertake a full systematic review, and for those reading and assessing the reliability of reviews to learn.

Wherever feasible, we advocate the use of systematic reviews, particularly where review outputs are intended for use in high-stakes decision making. However, whilst systematic reviews represent a gold standard in evidence synthesis, traditional reviews continue to be produced for many valid reasons, particularly by researchers or organizations with limited resources and competing demands and in instances where configurative and conceptual reviews are undertaken.

In situations where full systematic reviews are not an option, we argue that traditional reviews can serve important roles and that the value of traditional reviews can be considerably improved by applying the most contextually appropriate activities for increasing transparency, repeatability, and objectivity. This may be particularly beneficial for small- to medium-sized conservation organizations, students, and those working individually without the support of a large, active review group.

Acknowledgments

We thank 2 anonymous reviewers for helpful and constructive comments on an earlier draft. B.M. acknowledges a doctoral research grant from the Forest and Nature for Society (FONASO) Erasmus Mundus consortium. N.H. thanks MISTRA EviEM for allocation of time to editing of the manuscript. A.C. acknowledges a Policy Placement Fellowship funded by the Natural Environment Research Council, the UK Department for Environment Food and Rural Affairs and the Environment Agency.

Literature Cited

Bernes C, Bråthen KA, Forbes BC, Hofgaard A, Moen J, Speed JD. 2013*a*. What are the impacts of reindeer/caribou (*Rangifer tarandus L.*) on arctic and alpine vegetation? A systematic review protocol. Environmental Evidence **2**:http://www.environmentalevidencejournal.org/content/2/1/6.

Bernes C, Carpenter SR, Gårdmark A, Larsson P, Persson L, Skov C, Van Donk E. 2013*b*. What is the influence on water quality in temperate eutrophic lakes of a reduction of planktivorous and benthivorous fish? A systematic review protocol. Environmental Evidence **2**: http://www.environmentalevidencejournal.org/content/2/1/9.

Bilotta GS, Milner AM, Boyd IL. 2014. Quality assessment tools for evidence from environmental science. Environmental Evidence **3**: http://www.environmentalevidencejournal.org/content/3/1/14.

Borenstein M, Hedges LV, Higgins JPT, Rothstein H. 2009. Introduction to meta-analysis. Wiley, New York.

Civil Service. 2010. Rapid evidence assessment toolkit index. Civil Service, London. Available from http://www.civilservice.gov.uk/networks/gsystematic review/resources-and-guidance/rapid-evidence-assessment (accessed September 2014).

Collaboration for Environmental Evidence (CEE). 2013. Guidelines for systematic review and evidence synthesis in environmental management. Version 4.2. Environmental Evidence http://environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelines-version-4.2-final.pdf.

Collins A, Miller J, Coughlin D, Kirk S. 2014. The production of quick scoping reviews and rapid evidence assessments a how to guide. Beta version 2 (April 2014). Joint Water Evidence Group A, Defra, London. Available from https://connect.innovateuk.org/documents/3058188/3918930/The%20Production%20of%20QSRs%20and%20REAs-%20A%20How%20to%20guide.pdf?version=1.0 (accessed September 2014).

DfID (Department for International Development). 2014. How to guide: assessing the strength of evidence. Available from https://www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence (accessed September 2014).

Dickersin K. 1990. The existence of publication bias and risk factors for its occurrence. JAMA **263:**1385–1389.

Egger M, Smith GD. 1998. Bias in location and selection of studies. BMJ **316:**61–66.

Englund G, Sarnelle O, Cooper SD. 1999. The importance of data-selection criteria: meta-analyses of stream predation experiments. Ecology **80:**1132–1141.

Gates S. 2002. Review of methodology of quantitative reviews using meta-analysis in ecology. Journal of Animal Ecology **71:**547–557.

Gough D, Oliver S, Thomas J. 2012. An introduction to systematic reviews. Sage Publications, London.

Haddaway NR. 2015. A call for better reporting of conservation research data for use in meta-analyses. Conservation Biology DOI: 10.1111/cobi.12449.

Haddaway NR, Burden A, Evans CD, Healey JR, Jones DL, Dalrymple SE, Pullin AS. 2014*a*. Evaluating effects of land management on greenhouse gas fluxes and carbon balances in boreotemperate lowland peatland systems. Environmental Evidence **3**: http://www.environmentalevidencejournal.org/content/3/1/5

Haddaway NR, Styles D, Pullin AP. 2014*b*. Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: a systematic map. Environmental Evidence **3:**17. DOI: 10.1186/2047-2382-3-17.

Haddaway NR, Collins A, Coughlin D, Kirk S, Cooke S, Richards R, Stewart R, Johansson S, Knight TM. 2015. Proposal for establishment of CEE rapid evidence review methods. The Collaboration for Environmental Evidence: in press.

Halme P, Toivanen T, Honkanen M, Kotiaho JS, Mönkkönen M, Timonen J. 2010. Flawed meta-analysis of biodiversity effects of forest management. Conservation Biology **24:**1154–1156.

Harrison F. 2011. Getting started with meta-analysis. Methods in Ecology & Evolution **2:**1–10.

Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JAC. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ **343:**d5928.

Huntington BE. 2011. Confronting publication bias in marine reserve meta-analyses. Frontiers in Ecology and the Environment **9:**375–376.

Koricheva J, Gurevitch J. 2014. Uses and misuses of meta-analysis in plant ecology. Journal of Ecology **102:**828–844.

Koricheva J, Gurevitch J, Mengersen K. 2013. Handbook of meta-analysis in ecology and evolution. Princeton University Press, New Jersey.

Larsen PO, von Ins M. 2010. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics **84:**575–603.

Lortie CJ, Aarssen LW, Budden AE, Koricheva JK, Leimu R, Tregenza T. 2007. Publication bias and merit in ecology. Oikos **116:**1247–1253.

Macura B, Secco L, Pullin AS. 2013. Does the effectiveness of forest protected areas differ conditionally on their type of governance? Environmental Evidence **2** DOI: 10.1186/2047-2382-2-14.

McGowan J, Sampson M. 2005. Systematic reviews need systematic searchers. JMLA **93:**74–80.

Megill A. 2007. Historical knowledge, historical error. The University of Chicago Press, Chicago.

Michels C, Schmoch U. 2012. The growth of science and database coverage. Scientometrics **93:**831–846.

Møller AP, Jennions MD. 2001. Testing and adjusting for publication bias. Trends in Ecology & Evolution **16:**580–586.

Pautasso M. 2012. Publication growth in biological sub-fields: patterns, predictability and sustainability. Sustainability **4:**3234–3247.

Petticrew M. 2001. Systematic reviews from astronomy to zoology: myths and misconceptions. BMJ **322:**98–101.

Petticrew M, Roberts H. 2008. Systematic reviews in the social sciences: a practical guide. Blackwell Publishing, Oxford, Malden, MA.

Philibert A, Loyce C, Makowski D. 2012. Assessment of the quality of meta-analysis in agronomy. Agriculture, Ecosystems & Environment **148:**72–82.

Pullin AS, et al. 2013. Human well-being impacts of terrestrial protected areas. Environmental Evidence **2**: DOI:10.1186/2047-2382-2-19.

Pullin AS, Stewart GB. 2006. Guidelines for systematic review in conservation and environmental management. Conservation Biology **20:**1647–1656.

Randall NP, James KL. 2012. The effectiveness of integrated farm management, organic farming and agri-environment schemes for conserving biodiversity in temperate Europe – a systematic map. Environmental Evidence **1**: DOI: 10.1186/2047-2382-1-4.

Roberts PD, Stewart GB, Pullin AS. 2006. Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. Biological Conservation **132:**409–423.

Spencer L, Ritchie J, Lewis J, Dillon L. 2003. Quality in qualitative evaluation: a framework for assessing research evidence. Cabinet Office http://www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-38740.pdf (accessed September 2014).

Sterne JA, Egger M, Smith GD. 2001. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. BMJ **323:**101–105.

The Cochrane Collaboration. 2011. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]. Higgins J, Green S, editors. Available from www.cochrane-handbook.org (accessed September 2014).

Torgerson CJ. 2006. Publication bias: The Achilles' heel of systematic reviews? British Journal of Educational Studies **54:**89–102.

Woodcock P, Pullin AS, Kaiser MJ. 2014. Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: a methodology. Biological Conservation **176:** 54–62.