Environment International 114 (2018) 357–359

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Discussion

Meta-analysis is not an exact science: Call for guidance on quantitative synthesis decisions

Neal R. Haddaway[a,c,*], Trina Rytwinski[b]

[a] Mistra EviEM, Stockholm Environment Institute, Linnégatan 87D, Stockholm, Sweden
[b] Canadian Centre for Evidence-Based Conservation and Environmental Management, Institute of Environmental Sciences, Carleton University, Canada
[c] Africa Centre for Evidence, University of Johannesburg, Johannesburg, South Africa

ARTICLE INFO

Handling Editor: Paul Whaley

Keywords:
Quantitative synthesis
Meta-regression
Subgroup analysis
Evidence synthesis
Systematic reviewing
Evidence review

ABSTRACT

Meta-analysis is becoming increasingly popular in the field of ecology and environmental management. It increases the effective power of analyses relative to single studies, and allows researchers to investigate effect modifiers and sources of heterogeneity that could not be easily examined within single studies. Many systematic reviewers will set out to conduct a meta-analysis as part of their synthesis, but meta-analysis requires a niche set of skills that are not widely held by the environmental research community. Each step in the process of carrying out a meta-analysis requires decisions that have both scientific and statistical implications. Reviewers are likely to be faced with a plethora of decisions over which effect size to choose, how to calculate variances, and how to build statistical models. Some of these decisions may be simple based on appropriateness of the options. At other times, reviewers must choose between equally valid approaches given the information available to them. This presents a significant problem when reviewers are attempting to conduct a reliable synthesis, such as a systematic review, where subjectivity is minimised and all decisions are documented and justified transparently. We propose three urgent, necessary developments within the evidence synthesis community. Firstly, we call on quantitative synthesis experts to improve guidance on how to prepare data for quantitative synthesis, providing explicit detail to support systematic reviewers. Secondly, we call on journal editors and evidence synthesis coordinating bodies (e.g. CEE) to ensure that quantitative synthesis methods are adequately reported in a transparent and repeatable manner in published systematic reviews. Finally, where faced with two or more broadly equally valid alternative methods or actions, reviewers should conduct multiple analyses, presenting all options, and discussing the implications of the different analytical approaches. We believe it is vital to tackle the possible subjectivity in quantitative synthesis described herein to ensure that the extensive efforts expended in producing systematic reviews and other evidence synthesis products is not wasted because of a lack of rigour or reliability in the final synthesis step.

1. Background

Meta-analysis is the process by which the numerical results from across multiple studies are statistically combined, providing estimates of the overall mean effect and the variability around this mean (Smith and Glass, 1977). Such quantitative synthesis of study findings increases the effective power of analyses relative to single studies, and allows researchers to investigate effect modifiers and sources of heterogeneity that could not be easily examined within single studies; for example, the influence of political context on the effectiveness of a management intervention (Stewart, 2010). Since its development in the 1970s, meta-analysis has become increasingly popular for combining large bodies of evidence (Stewart, 2010; Haddaway, 2015).

Systematic review is a transparent, comprehensive and reliable means of summarising a body of evidence on a particular topic (Higgins and Green, 2011), and these evidence synthesis methods, too, are becoming increasingly popular (Haddaway et al., 2015) since their development in the 1990s and translation to environmental research in the mid 2000s (Pullin and Stewart, 2006). Many systematic reviewers will conduct a meta-analysis as part of their synthesis, but meta-analysis requires a niche set of skills that are not widely held by the environmental research community. Reviewers must therefore strive to include an expert on their review team or develop appropriate expertise within the team. External support is available through personal networks and contracting in experts for the latter stages of a review, but this is largely dependent on 'who you know', having a sufficient budget to cover

* Corresponding author.
E-mail addresses: neal_haddaway@hotmail.com, neal.haddaway@sei-international.org (N.R. Haddaway).

https://doi.org/10.1016/j.envint.2018.02.018
Received 24 November 2017; Received in revised form 7 February 2018; Accepted 10 February 2018
Available online 23 February 2018

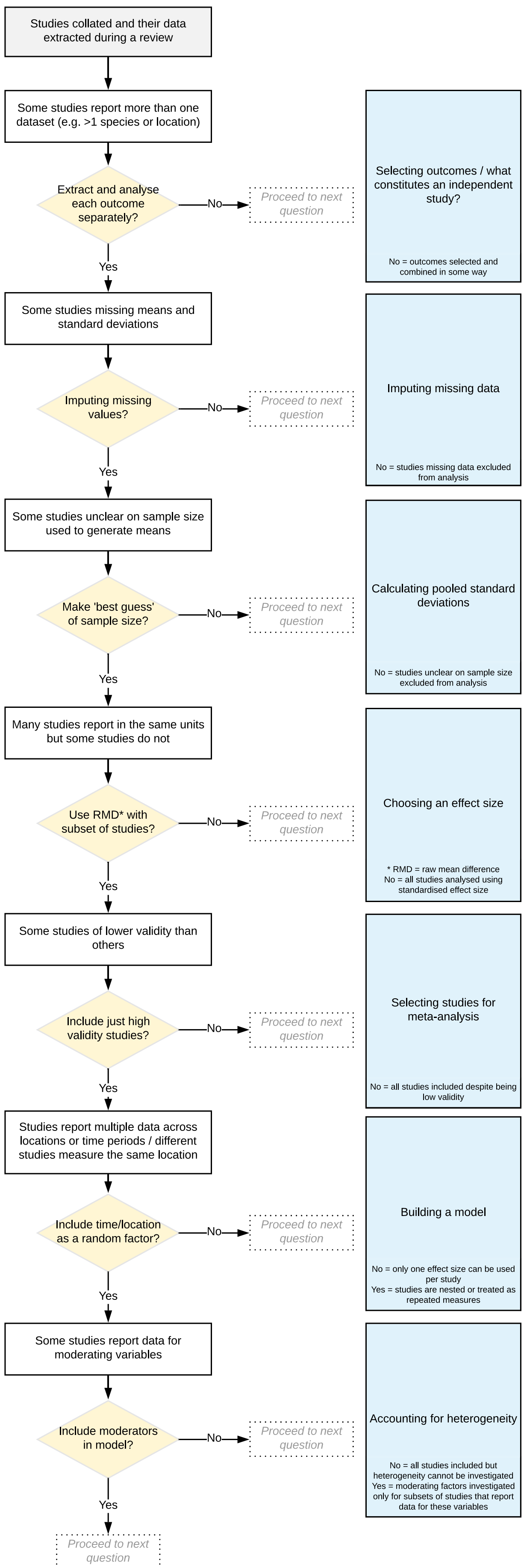

Fig. 1. Flow diagram illustrating examples of subjective decisions inherent in quantitative synthesis. The list is not exhaustive: many further questions could be encountered in data selection, extraction, preparation and modelling.

external staff costs, and in some cases (i.e., commissioned systematic reviews) having a flexible timeline that allows the additional time required to do so to be planned in from the outset.

2. The problem

Meta-analysis (and statistics in general) is not an exact science. Each step in the process of carrying out a meta-analysis requires decisions that have both scientific and statistical implications (see Fig. 1 for an illustrative example). One of the first and most critical decisions when synthesising environmental research is matching review questions with appropriate effect size metrics (Osenberg et al., 1999). Depending on the type of data, it is possible that only one measure is appropriate. For example, where all studies report greenhouse gas emissions in the same units, a raw mean effect size (treatment minus control) may be the obvious choice of metric, since the summary effect estimate will be inherently more meaningful than a standardized (and unitless) metric. However, in other cases, multiple effect sizes may be available (e.g., log response ratio or standardized mean difference), leaving the choice of which measure to use up to the reviewers (Rosenberg et al., 2013). Once an effect size measure is chosen, there are further decisions to face regarding the handling of the data from each primary study and the effect size calculation. These include: What constitutes an independent study? What should be considered a true replicate? How should outcomes be grouped, if at all? If multiple time points are presented for a treatment effect, should all time points be used in the calculation or just the most recent? Which data sets are methodologically sound enough to be included? In addition to these issues, the variances of effect size estimates can be computed in a variety of ways, and this requires an understanding of the sampling distribution of the chosen effect size measure. While many effect size measures have well known sampling distributions (e.g., *z*-transform correlation coefficient, standardized mean difference, etc), others do not, and in such cases an appropriate distribution may be derived in order to compute these variances (Gurevitch and Hedges, 1999). To derive a distribution for some such cases, one could for example identify parallels between the effect of interest and an effect with known statistical properties (Mengersen and Gurevitch, 2013). With these numerous decisions to be made by reviewers, there are times when a clear choice is either not obvious, or reviewers do not have a clear awareness of the statistical properties and underlying assumptions to make informed decisions. Of particular concern is that these decisions could ultimately impact the validity of statistical tests and confidence intervals in the analysis. Whilst there is a substantial body of guidance on how to conduct a meta-analysis and what choices to take (Page et al., 2017), there is limited guidance on what to do when a clear choice is not obvious.

Reviewers also face a number of decisions in addition to the choices described above – many of which involve considerable subjective judgement – with respect to model selection, and statistical approaches for analysis and inference (Thompson, 1994). These judgements include for example, decisions about the selection of a model (fixed-effect or random-effects), methods of analysis (maximum likelihood versus restricted maximum likelihood inference), whether studies should be grouped for subgroup analyses, and how to do so appropriately [i.e., being mindful of the potential for limited statistical power (Type I errors) and spurious significance with multiple analyses of numerous subgroups (Type II errors)], and whether outliers should be removed. Some decisions may be clear depending on the data and the context of analysis (for example fixed versus random effects meta-analysis), whilst other decisions may reflect more 'grey' issues. Whilst reviewers should – as recommended by Collaboration for Environmental Evidence (2018) –

state explicitly the rationales behind each of these decisions to minimise bias and increase transparency, there are also instances where one approach is as equally valid as another. These different (yet equally appropriate) analytical approaches may lead to different interpretations and implications of the statistical results. Thus, even with appropriate guidance on how to prepare data for meta-analysis and conduct the quantitative synthesis (of which there is currently a paucity, but see general introductory guidance in Borenstein et al., 2009; Koricheva et al., 2013), different reviews could potential find contrasting results despite using robust, valid approaches.

3. The solution

We propose three urgent, necessary developments within the evidence synthesis community. Firstly, we call on quantitative synthesis experts to improve guidance on how to prepare data for quantitative synthesis, providing explicit detail to support systematic reviewers. We believe that this guidance should account for the fact that there is a growing need for reviewers to conduct meta-analysis without having extensive statistical expertise. We suggest that this guidance should include decision-trees to practically support reviewers, and provide examples of poor practices that should be avoided. We believe that the guidance should be clear when certain practices are ‘best practice’, and when they are ‘the only practice’ that will yield a valid result.

Secondly, we call on journal editors and evidence synthesis coordinating bodies, such as the Collaboration for Environmental Evidence (http://www.environmentalevidence.org/) and Cochrane (http://www.cochrane.org/), to ensure that quantitative synthesis methods are adequately reported in a transparent and repeatable manner in published systematic reviews. In addition, these bodies should ensure that all potentially subjective decisions relating to quantitative synthesis are clearly justified in all reviews. In order to do this, we believe these actors must make a concerted effort to identify ‘weak points’ in quantitative synthesis methods where decisions are not clearly objective.

Thirdly, we provide a suggestion for systematic reviewers and meta-analysists wishing to conduct a reliable quantitative synthesis. We propose that, where faced with two or more broadly equally valid alternative methods or actions, reviewers conduct multiple analyses, presenting all options, and discussing the implications of the different analytical approaches. Where the results of these sensitivity analyses conflict or differ and there is not one clearly more appropriate analysis, reviewers should present results within a range of uncertainty. In all cases, transparency in the methods and results of quantitative syntheses is paramount, as is full justification of any choices made. We are not advocating for performing sensitivity analyses where certain methods are long-established as inappropriate, for example using a fixed effects meta-analysis where the populations of each study are known to be different, or where heterogeneity is known to be present a priori.

Finally, we call for review authors conducting quantitative synthesis to ensure that their raw and summarised data (i.e. not only the data collected but also the data entered into models following manipulation) and code are made freely available for reviewers, editors, and readers to repeat statistical analyses and understand in full all choices made during data-preparation and analysis.

We believe it is vital to tackle the possible subjectivity in quantitative synthesis described herein to ensure that the extensive efforts expended in producing systematic reviews and other evidence synthesis products is not wasted because of a lack of rigour or reliability in the final synthesis step (Chalmers and Glasziou, 2009). We believe our proposed calls to action and suggested sensitivity analysis approach to choices in data preparation and meta-analysis offer a solution to these problems. Urgent action from the evidence synthesis community is the only way to ensure these issues are properly dealt with.

Author contributions

Both authors developed the concept behind the manuscript, drafted the text and edited the final version. Both authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

The authors declare that they have no conflicts of interest.

References

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2009. Introduction to Meta-Analysis. John Wiley & Sons, Ltd, West Sussex, UK.
Chalmers, I., Glasziou, P., 2009. Avoidable waste in the production and reporting of research evidence. Obstet. Gynecol. 114, 1341–1345.
Collaboration for Environmental Evidence, 2018. Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0. www.environmentalevidence.org/information-for-authors [01/01/18].
Gurevitch, J., Hedges, L.V., 1999. Statistical issues in ecological meta-analyses. Ecology 80, 1142–1149.
Haddaway, N.R., 2015. A call for better reporting of conservation research data for use in meta-analyses. Conserv. Biol. 29, 1242–1245.
Haddaway, N., Woodcock, P., Macura, B., Collins, A., 2015. Making literature reviews more reliable through application of lessons from systematic reviews. Conserv. Biol. 29, 1596–1605.
Higgins, J.P., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons.
Koricheva, J., Gurevitch, J., Mengersen, K., 2013. Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press.
Mengersen, K., Gurevitch, J., 2013. Using Other Metrics of Effect Size in Meta-Analysis. In: Koricheva, J., Gurevitch, J., Mengersen, K. (Eds.), Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press, Princeton, USA, pp. 72–85.
Osenberg, C.W., Sarnelle, O., Cooper, S.D., Holt, R.D., 1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80, 1105–1117.
Page, M.J., Altman, D.G., McKenzie, J.E., Shamseer, L., Ahmadzai, N., Wolfe, D., Yazdi, F., Catalá-López, F., Tricco, A.C., Moher, D., 2017. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. J. Clin. Epidemiol. 95, 7–18.
Pullin, A.S., Stewart, G.B., 2006. Guidelines for systematic review in conservation and environmental management. Conserv. Biol. 20, 1647–1656.
Rosenberg, M.S., Rothstein, H.R., Gurevitch, J., 2013. Effect sizes: conventional choices and calculations. In: Koricheva, J., Gurevitch, J., Mengersen, K. (Eds.), Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press, Princeton, USA, pp. 61–71.
Smith, M.L., Glass, G.V., 1977. Meta-analysis of psychotherapy outcome studies. Am. Psychol. 32, 752.
Stewart, G., 2010. Meta-analysis in applied ecology. Biol. Lett. 6, 78–81.
Thompson, S.G., 1994. Why sources of heterogeneity in meta-analysis should be investigated. Br. Med. J. 309, 1351.