Meta-Analysis and Conservation Science

Karen Kovaka

(draft, do not circulate or cite without permission)

Abstract: Philosophical work on meta-analysis occurs exclusively in the context of biomedical research and revolves around a single question: Is meta-analytic evidence the best kind of evidence? I contribute to the epistemology of meta-analysis by identifying distinctive questions and puzzles that arise for meta-analysis in the context of conservation science, and I argue that philosophers should broaden their lens for studying this fascinating research method.

1. Introduction. Say you have an empirical question. Perhaps you want to know what the relationship between species richness and species productivity looks like, which are the best techniques for restoring damaged ecosystems, or the magnitude of the placebo effect. So you look to the scientific literature for answers.

For such questions, it can often seem, paradoxically, that science has generated both too much and not enough evidence. Too much, because there are thousands of published studies on each topic. Not enough, because none of these studies is, in itself, definitive. No token set of measurements is guaranteed to capture the generality you desire. Worse, some studies disagree with one another, and the experts themselves disagree about which results are the most reliable.

In such situations, a method called meta-analysis comes to the rescue. Meta-analysis is a set of statistical techniques for combining and analyzing data from multiple already-published studies. Its purpose is to provide an objective assessment of what conclusion(s) some complex body of evidence supports, and of the overall strength of the evidence.

Does meta-analysis succeed in its purpose? Well, the scientific consensus seems to be that meta-analysis is not merely *a* technique for combining evidence, but that the results of high-quality meta-analysis studies are *the* strongest and best form of scientific evidence available to us (Borenstein et al. 2021). But philosophers, predictably, have complicated the matter. There is an ongoing philosophical debate about the superiority of meta-analytic evidence in which one side argues that the choices and judgments researchers must make in order to conduct a meta-analysis compromise the method's objectivity (Stegenga 2011, 2018).

An idiosyncratic feature of this debate is that it only considers meta-analysis in the biomedical sciences. This is not so surprising, because meta-analysis originated in the biomedical context in the 1970s (Glass 1976), and its migration and application to other fields is more recent. The first ecological meta-analysis, for example, was published in 1991 (Jarvinen 1991). My goal here is to demonstrate that there are more interesting philosophical questions about meta-analysis than the narrow focus on biomedical science allows us to see. In this paper, I explore meta-analysis in a different domain—conservation science—showing that it has its own distinctive epistemic profile. Specifically, conservation meta-analysis typically occurs in a different decision-making context than biomedical meta-analysis, uses different criteria for selecting primary studies, and accepts greater heterogeneity. These differences raise a fascinating set of questions for the epistemology of meta-analysis and, I will argue, have implications for the debate about the superiority of meta-analytic evidence.

2. Meta-Analysis: A Primer. While the statistics of meta-analysis are complex, the overall logic of the method is quite accessible and worth a brief review (my discussion here draws heavily on Nakagawa and Santos 2012, Gurevitch et al. 2018, and Borenstein et al. 2021). A meta-analysis begins with a question, e.g., What does the species richness-species productivity relationship look like? Next, researchers search the published literature for studies which have produced data relevant to the question. If they determine there are enough studies of sufficient quality for a meta-analysis, they decide which studies from their initial search to include in the meta-analysis, and then the statistical work begins:

1. Researchers convert the outcome of each study to an *effect size*, a measurement suitable for combining with other studies.
2. Researchers combine the effect sizes for each study in a statistical model that accounts for differences in sample size across studies. The model outputs:
    a. a common effect size (or distribution of effect sizes) known as the *meta-analytic mean*, (e.g. a value or distribution of values for the species richness-species productivity relationship over all studies included in the analysis), and
    b. a measurement of *heterogeneity*, which is variation in study outcomes due to something other than chance (e.g. due to differences in study population or study design).

This much is common to meta-analyses in both biomedical science and conservation science. I shift now to identifying typical methodological differences between meta-analyses in these two domains. They fall into four categories: differences in goal, number of studies included in an analysis, choice of effect size statistic, and choice of statistical model.

- *Goal*: In biomedical science, most meta-analyses aim to quantify the effectiveness of a particular medical intervention (e.g. a vaccine) in a particular context, specifically the context of a clinical trial. Some conservation meta-analyses have similar goals, for example, to determine how effective a specific herbicide is at controlling a specific invasive species in a specific place. But more commonly, conservation meta-analyses aim to provide large-scale overviews of phenomena such as the relationship between species richness and species productivity or the overall effectiveness of certain restoration techniques across different environments.
- *Number of studies*: The difference in goal leads to a difference in the number of primary studies included in meta-analyses. A biomedical meta-analysis is likely to include fewer than 25 studies, while a conservation meta-analysis may include hundreds (Lau et al. 2013). One reason is that a meta-analysis aimed at testing the effectiveness of an intervention needs to make sure that the studies it includes are very similar to one another, while diversity among primary studies can be a virtue when the goal is to understand a phenomenon over a variety of populations, scales, and gradients.
- *Effect size statistic*: There are six common effect size statistics in meta-analysis (Borenstein et al. 2021), and the distribution of these statistics differs between biomedical and biological meta-analyses. A statistic called the *response ratio*, which is the natural log-proportional change in the means of a treatment and control group (Lajeunesse 2011), is the most common choice in ecology and evolution meta-analyses, but it is rarely used in other domains (Borenstein et al. 2021; Nakagawa and Santos 2012).
- *Statistical model*: There are two basic types of statistical model in meta-analysis. A *fixed-effect model* (Hedges and Olkin 1985) assumes that there is one true effect size,

> approximated by the meta-analytic mean, and that differences in measured effect size across studies are due to sampling variance. By contrast, a *random-effects model* assumes that there is a different true effect size for each study, due to underlying in differences in study population, conditions, design, etc. The results of a random-effects meta-analysis are in the form of a distribution of effect sizes, "situationally correct answers from an imagined universe of individual studies that might have been performed" (Imrey 2020, p. 1). Random-effects models are relatively more common in conservation science than in biomedical science, and in fact, the rapid acceptance of random-effects models is a "methodological innovation" credited to biological meta-analysis (Gurevitch et al. 2018, p. 179).

Now, methodological differences do not always indicate interesting underlying epistemic differences, but in this case, they do. Attention to variation in how meta-analyses are conducted in biomedicine and conservation leads us to interesting epistemic questions that are harder to see from the perspective of biomedical meta-analysis alone.

3. The Decision-Making Context. The first feature of the distinctive epistemic profile of conservation meta-analysis is its decision-making context. By decision-making context, I mean what decisions take meta-analysis as their input, and what standards govern those decisions. Two key differences in decision-making context lead to meta-analysis playing a different epistemic role in conservation than in biomedicine.

First, typical decisions in biomedicine that take meta-analysis as their input are decisions about whether to approve new medical interventions. Published discussions of meta-analyses express the goal of these studies in terms of their potential to inform such decisions: “Meta-analyses are conducted to assess the strength of evidence present on a disease and treatment. One aim is to determine whether an effect exists; another aim is to determine whether the effect is positive or negative…” (Haidich 2010, p. 30). Once a medical intervention is approved, clinicians may look to meta-analysis results when deciding whether to prescribe it in a particular case, but this is not the primary or intended use of biomedical meta-analyses. The typical decision that takes biomedical meta-analysis as its input is, “Should this intervention be approved?” and not, “Should I use this intervention in my case?”

In conservation, by contrast, no formal approval process for interventions exists. The decisions which take meta-analysis as their input are, as a result, much more likely to be decisions about which of a menu interventions or management strategies to use in particular cases: “Conservation management involves day-to-day decision making by a wide range of individuals…All face decisions regarding what actions they should take to achieve objectives and most will involve a level of uncertainty of outcome” (Pullin et al. 2004, p. 245).

Second, decisions about whether to approve a new medical intervention are governed by a strict evidence of effectiveness requirement, as codified in the Federal Food, Drug and Cosmetic Act of 1962. This requirement is a standardized version of the ancient Hippocratic prescription to do no harm. This means that a regulatory body with the authority to approve the therapeutic use of a new drug will take no action with respect to the disease the drug is potentially effective against until researchers show sufficient evidence of effectiveness.

There is no such constraint on conservation decisions about the use of interventions. The lack of constraint means that in practice, few conservation decisions actually take meta-analyses or other forms of systematic review as inputs (Sutherland and Wordley 2017). Advocates of evidence-based conservation have long lamented this reality and called for conservation decision-making to take evidence-based medicine as a model (e.g. Pullin and Stewart 2006, Sutherland 2004). But what they mean by this is not that conservation should adopt a similar evidence of effectiveness standard to the one used in biomedicine. In fact, they argue that "…conservation cannot adopt evidence frameworks, tools, and guidance from other fields wholesale, since the needed and available evidence as well as the standards for evidence quality vary vastly across disciplines" (Salafsky et al. 2019, p. 2).

It is understood that conservation decision-makers are not in a position to refrain from implementing interventions or management strategies unless a specific threshold of evidence of effectiveness is met. Decision-makers cannot put choices about which biodiversity conservation or ecosystem restoration actions to take on hold until an ideal amount of evidence is available. The necessity of action in the face of uncertainty is a hallmark of conservation decision-making. So, while advocates of evidence-based conservation call for incorporating meta-analysis into decision-making, they do not suggest that meta-analysis should establish a particular level of evidence of effectiveness for an intervention before decision-makers consider using it.

So, meta-analysis plays different epistemic roles in conservation and biomedicine. It supports decision-making in both cases, but the decisions it informs are different. In biomedicine the ideal is for a meta-analysis to establish the effectiveness of any intervention before it is implemented in practice, while in conservation it is understood that decision-makers must sometimes, if not often, implement interventions in the absence of this level of evidence of effectiveness. Meta-analysis has more of a gatekeeping function in biomedicine, which means that key epistemic questions in this domain are whether meta-analysis studies really are the best kind of evidence for the effectiveness of medical interventions, and whether the evidence from these studies really is sufficient for determining effectiveness. But in conservation, where the ideal function for meta-analysis is to inform a myriad of context-specific decisions, the key epistemic questions are about how to understand the implications meta-analyses have for this variety of contexts. The next two sections explore these key epistemic questions for conservation meta-analysis.

4. All the Evidence Versus the Best Evidence. A second feature of the distinctive epistemic profile of conservation meta-analysis is how inclusive it is about selecting primary studies to include in an analysis. The standard in biomedicine is to include only the best evidence, that is, evidence from randomized control trials (Higgins 2019). This is an additional reason why biomedical meta-analyses tend to include fewer primary studies than conservation meta-analyses. One worry about this standard is that it violates the principle of total evidence (Stegenga 2018), while a response is that a historical analysis of science shows we may be pragmatically justified in excluding data, because doing so allows the scientific community to reach the right answer to a question more quickly than considering all of the data would (Holman 2019).

In any case, conservation meta-analyses have more relaxed standards for including primary studies. Randomized control trials, which randomly assign experimental units (e.g. human participants or agricultural plots) to control and treatment groups, are not only rare in conservation science, they are often impossible (Pynegar et al. 2021). A study of an actual effort to restore an abandoned strip mine site, for example, cannot be a randomized control trial, because there is only one experimental unit. Even studies with multiple experimental units may not have sufficient sample sizes for a randomized control trial, or random assignment may not be feasible. So conservation meta-analyses are pushed to include primary studies that use at least some methods other than randomized control trials.

But which methods? One could limit the list of appropriate experimental methods to studies that at least use control and treatment groups, and measure both groups before and after an intervention (known as *before-after-control-impact*), but drop the randomness requirement. Or, one could allow both simple control-impact and simple before-after designs, but exclude studies that merely monitor. Unfortunately, what are considered the most reliable study designs (e.g. before-after-control-impact) are also the rarest, while simple monitoring studies are the most common (Christie et al. 2020).

In light of these facts, conservation meta-analyses often include studies with all of the above designs. As an example, consider a study by Jones et al. (2018) that examined ecological restoration of damaged ecosystems. This meta-analysis measured the rate and extent of recovery after large-scale human disturbance (e.g. mining, logging, oil spills) in different ecosystems. It included 400 primary studies and over 5000 variables. But only 49 of these variables used control-impact-before-after designs.

The worry about this approach is that including studies of worse design compromises the reliability of meta-analytic results. Those compelled by this worry argue that conservation meta-analysis should be more like biomedical meta-analysis: it should use stringent criteria for including primary studies, even if this means meta-analysis becomes impossible for many questions because not enough primary studies meet the criteria (Whittaker 2010). But the much more common view in conservation is that rather than using stringent selection criteria to reject most potential studies outright, the better approach is to empirically test for differences in study design that actually make a difference to the outcome of the meta-analysis:

> That is, to gather all the studies relevant to the conceptual topic under study, and then empirically test whether these differences (i.e., any factor presumably affecting quality) actually influence research outcomes. For example, contrasting the findings from groups of studies with and without these problems, or through sensitivity analyses where collections of studies are excluded from the overall synthesis to evaluate their weight on the pooled conclusions. Should a meta-analysis detect a difference between these groups, then (1) this provides practical information for future experiments to avoid these problems, (2) there is a solid rationale for why these studies should be included or excluded from the overall review, and (3) more sophisticated approaches…can be used to integrate issues on quality into the overall analysis (Lajeunesse 2010).

I will not take a position on how stringent selection criteria should be. Instead, the takeaway is that the widespread practice of including many studies with differing designs raises distinctive epistemic issues, which are not present in biomedical meta-analysis. How good is sensitivity analysis at identifying studies and study designs which, due to their poor quality, influence meta-analytic results? How well does the inclusive approach to primary study selection serve the goal of identifying broad patterns and reaching coarse generalizations? For what kinds of research questions and under what circumstances is it preferable to include more, rather than fewer, data in a meta-analysis?

5. Heterogeneity. Deeply connected to the question of selection criteria for primary studies in meta-analysis is the issue of heterogeneity. Study design is only one dimension of the selection criteria question. Another dimension has to do with how similar primary studies should be in other ways, e.g. in terms of their focal population, spatial scale, environmental setting, etc. All of these are sources of variation that can influence the outcome of a study, so including studies that vary along these dimensions in a meta-analysis of the effectiveness of, say, different ecological restoration techniques, introduces the possibility the meta-analytic results will reflect variation in the primary studies rather than the actual effectiveness of the restoration techniques in question. For this reason, *heterogeneity* is the enemy in biomedical meta-analysis. As the Cochrane Handbook states, “Meta-analysis should only be considered when a group of studies is sufficiently homogeneous in terms of participants, interventions and outcomes” (Higgins 2019, 9.5.1).

But whereas in biomedicine there is a wealth of published studies that are arguably homogenous, this is rarely the case in conservation science. The universe of targets for an intervention in conservation includes many different species, biomes, temporal and spatial scales, and approaches to implementing the same intervention, to name just a few common sources of heterogeneity. Thus, there is great potential for differences in effect size statistic from primary study to primary study to be partly or largely due to heterogeneity. What to do?

Methodologically, there are many statistical resources for identifying and quantifying heterogeneity. So it’s not so much a concern, at least not an in-principle one, that researchers or decision-makers will draw illegitimate conclusions from heterogeneous studies where the heterogeneity goes undetected. To be sure, there are quite a few poor-quality conservation meta-analyses, but given that the resources to overcome this problem exist, it’s not particularly interesting to focus on.

A more interesting set of problems has to do with translating the results of heterogeneous studies to conservation science’s decision-making context. In an important sense, heterogeneity is a virtue in conservation meta-analysis:

> …when the goal is to reach broad generalizations, the population of studies may be large and heterogeneous and, although estimating the main effect of a particular phenomenon or experimental treatment may be important, identifying sources of heterogeneity in outcomes is often central to understanding the overall phenomenon. Meta-analyses undertaken with the aim of reaching broad generalizations deliberately incorporate results from heterogeneous populations so

> that broad generalizations and the factors that modify them can be examined and tested (Gurevitch et al 2018, p. 177).

Yet, when we think about a group of land managers tasked with restoring a stream, or a state ecologist deciding whether a controlled burn is a good measure for controlling an invasive plant, broad generalizations seem less useful: "…while meta-analysis can characterize broad patterns, by definition, it operates at a coarse resolution. By contrast, restoration is highly context-dependent and practitioners make informed decisions based on project-specific factors" (Larkin et al. 2019, p. 3).

More concretely, a meta-analysis with high heterogeneity must use a random-effects model. The confidence interval for the results of a random-effects meta-analysis is wider than that for a fixed-effect meta-analysis. Recall that the meta-analytic mean a random-effects model reports is best understood as the "center of a dispersed range" of effect sizes (Imrey 2020, p. 2). The standard deviation around this mean can be quite high. In an example focused on a global meta-analysis of burnout prevalence in medical professionals, Imrey (2020) shows that given the burnout prevalent estimate of 32% and the standard deviation of 15%, "…true target burnout prevalences for one-third of studies would be outside the range of 17% to 47% and, for 10%, outside the range of 7% to 57%. Thus, the 32% average provides only very modest guidance on what to expect for any given situation because the studies are so heterogeneous" (p. 2). The force of this point remains even after applying statistical tools such as subgroup analysis and meta-regression for understanding the sources of heterogeneity.

Heterogeneity thus raises fascinating questions about meta-analysis and conservation decision-making, as do novel attempts to address these questions within the framework of meta-analysis. Consider Shackelford et al.'s (2021) development of *dynamic meta-analysis*. Their aim is to make global data about conservation questions relevant to specific contexts. Recognizing that we are unlikely to ever have homogenous meta-analyses for specific types conservation decisions, and definitely not on the timescale required to response to global environmental crisis, Shackelford hypothesize that a web application which assembles a variety of studies relevant to a particular research question or intervention, and which also lets practitioners filter and weight studies according to their particular context, can help to address this challenge. They create a proof-of-concept tool at www.metadataset.com where users can perform this kind of analysis on sets of studies on two topics: invasive species and cover crops.

6. How Good Is Meta-Analytic Evidence? To summarize: conservation meta-analysis has a distinctive epistemic profile. A desideratum for conservation meta-analyses is that they can be used to inform decisions where some action must be taken, and the priority is on understanding how the available evidence comes to bear on specific contexts, rather than on determining whether the evidence meets a threshold for action. The meta-analyses that can possible inform conservation decisions have a range of goals—sometimes to test the effectiveness of particular intervention in a particular place, but often not—and are typically heterogeneous and inclusive of all or much of the available evidence. These facts suggest that the body of evidence conservation meta-analysis provides is qualitatively different than the body of evidence biomedical meta-analysis provides, and the best practices for incorporating this evidence into decisions will also be different.

My primary goal has been to make the case for broadening the epistemology of meta-analysis. The questions I have raised so far are interesting in their own right, and worthy of philosophical attention. They are also relevant to the existing philosophical literature on meta-analysis, which I turn to now.

Briefly, the existing philosophical literature on meta-analysis centers on whether meta-analysis provides better evidence about the effectiveness of medical interventions than other methods do. The critical position, represented by Stegenga (2011, 2018), is that meta-analysis suffers from ineliminable arbitrariness: researchers must make a variety of choices in the course of conducting a study, and these choices can be made in different ways, which can and do lead to conflicting results. These include choices about which primary studies to include and how to statistically analyze them. In response to Stegenga, Holman (2019) argues that over time it is possible to adjudicate between the different choices which lead to conflicting results, and thus to resolve these conflicts. Widening the lens to include conservation meta-analysis in addition to biomedical meta-analysis suggests two takeaways for this debate.

First, some of the arguments in this debate rely on assumptions about meta-analysis that are specific to biomedicine. One of Stegenga's (2018) concerns about meta-analysis is that, due to heterogeneity, both among studies and between study and clinical populations, extrapolation from the research to the clinical context is not straightforward, yet biomedical experts use a simplistic heuristic of assuming that extrapolation is legitimate, unless there is a clear reason to think otherwise. This may be the case in biomedicine, but in conservation science, the reality of heterogeneity is very much appreciated, and simple extrapolation is not the norm.

Conservation meta-analysis also differs from biomedical meta-analysis with respect to the influence of industry. What philosophers call "industrial selection" (Holman and Bruner 2017) is a serious problem in biomedical research. Briefly, it is possible for the pharmaceutical industry to bias drug trials toward finding evidence of effectiveness, to suppress "inconvenient" research, and in other ways to skew the data that go into meta-analyses in ways that may make them systematically biased. In conservation, though industry may influence some aspects of research, such as agriculture, there are fewer resources dedicated to producing particular research results. This is not to say that other kinds of bias, such as publication bias, are not still concerning, but the landscape surrounding issues of bias in conservation is quite different, and conclusions about the reliability of meta-analysis that depend on accusations of industrial selection are not necessarily applicable to conservation.

Second, the primary critique of meta-analysis in biomedicine is that the many different choices researchers make in study design and implementation produces conflicting results, and that these choices cannot be rationally constrained in a way that solves the problem. This feature, which Stegenga calls *malleability*, is an even more serious problem in conservation than in biomedicine. Different standards for choosing research questions, managing primary study selection, and handling heterogeneity mean that researchers in conservation meta-analysis have even more choice points, which provide even more opportunities for conflicting results. Not only is this a greater problem in conservation science, researchers also have a greater sense of awareness of and humility about this problem. Conflicting results are expected, and even seen as an opportunity for identifying knowledge gaps, revealing knew information, and developing better studies in the future. In this way, conservation science supports Holman's call to view meta-analysis through a social epistemic lens. Meta-analysis generates puzzles and questions just as much as it answers them. It is part of an ongoing process of managing underdetermination, not a final response to it. Asking whether it produces the best evidence is ultimately less interesting than asking how it is able to iteratively improve the inferences it makes from a given body of data.

References

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R.K., Williams, D.R., Wordley, C.F.R, & Sutherland, W. J. (2021). The challenge of biased evidence in conservation. *Conservation Biology*, *35*(1), 249-262.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175-182.

Hedges L, Olkin I (1985) *Statistical methods for meta-analysis*. Academic Press, New York

Hedges LV, Vevea JL (1998) Fixed- and random-effects models in meta-analysis. *Psychol Methods 3:486–504*

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Holman, B. (2019). In defense of meta-analysis. *Synthese*, *196*(8), 3189-3211.

Holman, B., & Bruner, J. (2017). Experimentation by industrial selection. *Philosophy of Science*, *84*(5), 1008-1019.

Imrey, P. B. (2020). Limitations of meta-analyses of studies with high heterogeneity. *JAMA network open*, *3*(1), e1919325-e1919325.

Jarvinen, A. A meta-analytic study of the effects of female age on laying date and clutch size in the Great Tit *Parus major* and the Pied Flycatcher *Ficedula hypoleuca. Ibis* **133,** 62–67 (1991).

Jones, H. P., Jones Peter C., Barbier Edward B., Blackburn Ryan C., Rey Benayas Jose M., Holl Karen D., McCrackin, M., Meli, P., Montoya, D., & Mateos David Moreno. (2018). Restoration and repair of Earth's damaged ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20172577.

Lajeunesse, M. J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, *92*(11), 2049-2055.

Lajeunesse, M. J. (2010). Achieving synthesis with meta-analysis by combining and comparing all available studies. *Ecology*, *91*(9), 2561-2564.

Larkin, D. J., Buck, R. J., Fieberg, J., & Galatowitsch, S. M. (2019). Revisiting the benefits of active approaches for restoring damaged ecosystems. A Comment on Jones HP et al. 2018 Restoration and repair of Earth's damaged ecosystems. *Proceedings of the Royal Society B*, *286*(1907), 20182928.

Lau, J., Rothstein, H.R., & Stewart, G.B. 2013. "History and Progress of Meta-analysis." In Koricheva, J., Gurevitch, J., & Mengersen, K. (Eds.). (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.

Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*(5), 1253-1274.

Pullin, A. S., Knight, T. M., Stone, D. A., & Charman, K. (2004). Do conservation managers use scientific evidence to support their decision-making?. *Biological conservation*, *119*(2), 245-252.

Pullin, A. S., & Stewart, G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation biology*, *20*(6), 1647-1656.

Pynegar, E. L., Gibbons, J. M., Asquith, N. M., & Jones, J. P. (2021). What role should randomized control trials play in providing the evidence base for conservation?. *Oryx*, *55*(2), 235-244.

Salafsky, N., Boshoven, J., Burivalova, Z., Dubois, N. S., Gomez, A., Johnson, A., Aileen, L., Margoluis, R., Morrison, J., Muir, M., Pratt, S.C., Pullin, A.S., Salzer, D., Stewart, A., Sutherland, W.J., & Wordley, C. F. (2019). Defining and using evidence in conservation practice. *Conservation Science and Practice*, *1*(5), e27.

Shackelford, G. E., Martin, P. A., Hood, A. S., Christie, A. P., Kulinskaya, E., & Sutherland, W. J. (2021). Dynamic meta-analysis: a method of using global evidence for local decision making. *BMC biology*, *19*(1), 1-13.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational researcher*, *15*(9), 5-11.

Stegenga, J. (2018). *Medical nihilism*. Oxford University Press.

Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence?. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences*, *42*(4), 497-507.

Sutherland, W. J., & Wordley, C. F. (2017). Evidence complacency hampers conservation. *Nature ecology & evolution*, *1*(9), 1215-1216.

Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in ecology & evolution*, *19*(6), 305-308.

Whittaker, R. J. (2010). Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. *Ecology*, *91*(9), 2522-2533.