

SYMPOSIA PAPER

# Meta-Analysis and Conservation Science

Karen Kovaka 

Virginia Tech, Department of Philosophy, Major Williams Hall, 229, 220 Stanger St, Blacksburg, VA 24060  
Email: [kkovaka@vt.edu](mailto:kkovaka@vt.edu)

(Received 15 October 2021; revised 18 March 2022; accepted 14 May 2022; first published online 09 June 2022)

## Abstract

Philosophical work on meta-analysis focuses on biomedical research and revolves around the question: Is meta-analytic evidence the best kind of evidence? By focusing on conservation science rather than biomedical science, I identify further questions and puzzles for meta-analysis and show their importance for the epistemology of meta-analysis.

## 1. Introduction

Say you have an empirical question. Maybe you want to know how effective different ecological restoration strategies are, or how big the placebo effect is. If there is existing research on your question, you look to the scientific literature for answers.

For such questions, it can often seem 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 no study is definitive on its own. No token set of measurements captures the generality you desire. Worse, some studies disagree with one another, and experts disagree about which results are most reliable.

In these situations, a method called meta-analysis is supposed to come to the rescue. Meta-analysis is a set of statistical techniques for combining and analyzing data from existing studies. Its purpose is to assess what conclusions some body of evidence supports, and how strongly it supports those conclusions.

Does meta-analysis succeed in this purpose? The scientific consensus is that meta-analysis is not merely a technique for combining evidence, but the strongest and best technique for combining scientific evidence (Borenstein et al. 2021). But philosophers, predictably, have complicated the matter. In fact, much of the small philosophical literature on meta-analysis debates whether meta-analytic evidence really is the evidence (Stegenga 2011, 2018), while the rest engages with the method's conceptual foundations (Fletcher et al. 2019; Vieland and Chang 2019; Wüthrich and Steele 2019).

This literature focuses on meta-analysis in the biomedical sciences, and meta-analysis in other sciences gets much less attention. This is not surprising: meta-analysis originated in the biomedical context in the 1970s (Glass 1976), and its application to other fields is more recent. Yet there are important epistemological issues about meta-analysis that the narrow focus on biomedical science has made it difficult to see. In this paper, I consider meta-analysis in a domain that philosophers have

not addressed—conservation science. I draw three key contrasts between meta-analysis in the biomedical sciences and in conservation science: the role meta-analyses play in decision-making, the criteria for selecting which studies to include, and the degree of acceptance of statistical heterogeneity. Attention to these differences allows me to (a) identify questions that the epistemology of meta-analysis can and should explore, and (b) suggest a reframing of the philosophical debate about meta-analytic evidence.

## 2. Meta-Analysis: A Primer

While meta-analyses can be statistically complex, the overall logic of the method is quite accessible. A meta-analysis begins with a question. Here are two examples:

- How big is the placebo effect when patients know they are receiving a placebo (von Wernsdorff et al. 2021)?
- How much do ecosystems recover after large-scale, human-caused disturbances (Jones et al. 2018)?

Next, researchers search the published literature for studies relevant to the question. If they determine there are enough studies of sufficient quality for a meta-analysis, they decide which ones to include in the meta-analysis, and 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 each effect size 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*, and
  - b. a measurement of *heterogeneity*, which is variation in study outcomes due to something other than chance.

This much is common to meta-analyses in both biomedical science and conservation science. But let's explore our two examples in more detail in order to identify typical methodological differences between meta-analyses in these areas of science.

In the case of the placebo effect study, researchers screened 2,028 citations and selected 11 studies for the analysis. They used an effect size measure known as the standardized mean difference (the effect of the intervention relative to the variability of the study) and combined these effect sizes in a random-effects statistical model. They found a large overall effect (0.72) but also high heterogeneity (76%). If they excluded the 4 studies most likely to be biased, the overall effect fell to 0.49, with low heterogeneity (4%).

In the case of the ecological restoration study, researchers screened 972 citations and selected 400 for the analysis. They used an effect size measure known as a response ratio (the ratio of the actual effect to the goal effect for each study variable) to measure the extent of ecosystem recovery and combined these effect sizes in a general linear mixed model. They found positive but incomplete recovery rates in

all cases, and that the extent of recovery varied both with disturbance type (e.g. logging vs. oil spills) and ecosystem type (e.g. forest vs. wetland).

These examples highlight four key methodological contrasts: goals of the analysis, number of studies included in 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 in a particular context, most often a clinical trial. Some conservation meta-analyses have similar goals, for example, to determine how effective a specific herbicide is at controlling an invasive species in a particular place. But more commonly, conservation meta-analyses aim to provide large-scale overviews of phenomena such as 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). 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. A *random-effects model* assumes that there is a different true effect size for each study, due to underlying 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. Though some bio-medical studies do use random-effects models, they are more common in conservation science, and the rapid acceptance of random-effects models is a “methodological innovation” credited to biological meta-analysis (Gurevitch et al. 2018, 179).

Methodological differences do not always indicate interesting underlying epistemic differences, but in this case, they do. Conservation meta-analysis has a distinctive epistemic profile, which we will now explore.

### 3. The decision-making context

The first feature of the distinctive epistemic profile of conservation meta-analysis is its decision-making context. Two differences in decision-making context—differences in which decisions take meta-analysis as their input, and differences in the standards governing those decisions—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, 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 their primary use. 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, no formal approval process for interventions exists. The decisions which take meta-analysis as their input are much more likely to be decisions about which of a menu of interventions or management strategies to use in particular cases (Pullin et al. 2004, 245), that is, “Should I use this intervention in my case?”

Second, decisions about whether to approve a new medical intervention are governed by a strict evidence of effectiveness requirement. In the U.S., this is codified in the Federal Food, Drug and Cosmetic Act of 1962. This requirement, which is necessary but not sufficient for approval, 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 lament this reality and call for conservation decision-making to take evidence-based medicine as a model (e.g. Pullin and Stewart 2006). But they do not mean 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, 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. 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 biomedical decision making. First, the decisions it informs are different. Second, 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, perhaps often, implement interventions in the absence of this level of evidence of effectiveness. Meta-analysis has more of a gatekeeping function in biomedicine, so 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 its standard for selecting primary studies to include in an analysis. The standard in biomedicine is to include only what is considered the best evidence, which is evidence from randomized control trials (Higgins et al. 2019). This is one 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).

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. For example, the meta-analysis of the recovery of damaged

ecosystems from Section 2 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. (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 that 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 et al. 2019, 9.5.1).

But whereas in biomedicine there is a wealth of published studies that are arguably homogenous, or homogenous enough, 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 heterogeneity will go 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 ... (Gurevitch et al. 2018, 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, 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. The meta-analytic mean a random-effects model reports is best understood as the "center of a dispersed range" of effect sizes (Imrey 2020, 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" (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 homogeneous 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 created a proof-of-concept tool at [www.metadataset.com](http://www.metadataset.com) where users can perform this kind of analysis on sets of studies on two topics: invasive species and cover crops.

## 6. The value of meta-analytic evidence

What has our exploration of the distinctive epistemic profile of conservation meta-analysis revealed? A desideratum for conservation meta-analyses is that they can 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 possibly 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 that the best practices for drawing conclusions from the evidence and incorporating the evidence into decision-making are also different.

The issue of how exactly these best practices differ raises a series of novel questions for the epistemology of meta-analysis, which I have already highlighted. A further question, which I have not yet mentioned, is whether differences in methodology and practice in conservation meta-analyses ought to inform biomedical meta-analyses, or vice versa. I now turn to ways in which the distinctive profile of conservation meta-analysis is relevant to the philosophical debate about the quality of meta-analytic evidence.

This debate asks whether meta-analysis provides better evidence than other methods of amalgamating evidence 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. Attention to conservation meta-analysis suggests two take-aways for this debate.

First, some of the arguments in this debate rely on assumptions 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, the pharmaceutical industry can bias drug trials toward finding evidence of effectiveness, suppress “inconvenient” research, and otherwise skew the data that go into meta-analyses in ways that systematically bias the analyses. In conservation, though industry may influence some aspects of research (e.g. 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 across the board.

Second, the primary critique of meta-analysis in biomedicine is that the many choices researchers make in study design and implementation produce conflicting results, and that these choices cannot be rationally constrained in a way that solves the problem. This feature, known as *malleability* (Stegenga 2011), is a 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 more choice points, thus 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 new information, and developing better studies in the future.

These facts support Holman’s (2019) and Jukola’s (2015) 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. Further, the good of meta-analysis is not that it produces a once-and-for-all answer to a research question. It is better understood as a way of crystallizing the state of the evidence—in all its messiness—at a moment in time. This crystallization has a variety of values, not just one. This means that philosophers should not just ask whether meta-analytic evidence is the best evidence. We should also investigate the various epistemic benefits (and shortcomings) of meta-analysis, in order to develop a more complete picture of the contributions this method of evidence amalgamation can and does make across the sciences.

**Acknowledgements.** Thanks to Samuel Fletcher, Gil Hersch, Billy Monks, Wendy Parker, and Jacob Stegenga for their comments on this manuscript.

## References

- Borenstein, Michael, Larry Hedges, Julian Higgins, and Hannah Rothstein. 2021. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons.
- Christie, Alec. P., Tatsuya Amano, Philip A. Martin, Silviu O. Petrovan, Gorm E. Shackelford, Benno I. Simmons, Rebecca K. Smith, David R. Williams, Claire F. R. Wordley, and William J. Sutherland. 2020. “The Challenge of Biased Evidence in Conservation.” *Conservation Biology* 35(1):249–62.
- Fletcher, Samuel C., Jürgen Landes, and Roland Poellinger. 2019. “Evidence Amalgamation in the Sciences: An Introduction.” *Synthese* 196(8):3163–88.

- Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher* 5(10): 3–8.
- Gurevitch, Jessica, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. 2018. "Meta-Analysis and the Science of Research Synthesis." *Nature* 555(7695):175–82.
- Haidich, Anna-Bettina. 2010. "Meta-Analysis in Medical Research." *Hippokratia* 14(Suppl. 1):29–37.
- Hedges Larry, and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Higgins, Julian P., James Thomas, Jacqueline Chandler, Miranda Cumpston, Tian Jing Li, Matthew J. Page, and Vivian A. Welch, eds. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons.
- Holman, Bennett. 2019. "In Defense of Meta-Analysis." *Synthese* 196(8):3189–211.
- Holman, Bennett, and Justin Bruner. 2017. "Experimentation by Industrial Selection." *Philosophy of Science*, 84(5):1008–19.
- Imrey, Peter B. 2020. "Limitations of Meta-Analyses of Studies with High Heterogeneity." *JAMA Network Open* 3(1):e1919325–e1919325.
- Jones, Holly P., Peter C. Jones, Edward B. Barbier, Ryan C. Blackburn, Jose M. Rey Benayas, Karen D. Holl, Michelle McCrackin, Paul Meli, Daniel Montoya, & David Moreno Mateos. 2018. "Restoration and Repair of Earth's Damaged Ecosystems." *Proceedings of the Royal Society B: Biological Sciences*, 285(1873):20172577.
- Jukola, Saana. 2015. "Meta-Analysis, Ideals of Objectivity, and the Reliability of Medical Knowledge." *Science & Technology Studies* 28(3):101–21.
- Lajeunesse, Marc J. 2010. "Achieving Synthesis with Meta-Analysis by Combining and Comparing All Available Studies." *Ecology* 91(9):2561–64.
- Lajeunesse, Marc J. 2011. "On the Meta-Analysis of Response Ratios for Studies with Correlated and Multi-Group Designs." *Ecology* 92(11):2049–55.
- Larkin, Daniel J., Robert J. Buck, John Fieberg, and Susan M. Galatowitsch. 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, Joseph, Hannah R. Rothstein, and Gavin B. Stewart. 2013. "History and Progress of Meta-Analysis." In *Handbook of Meta-Analysis in Ecology and Evolution*, edited by Julia Koricheva, Jessica Gurevitch, and Kerrie Mengersen, 407–419. Princeton: Princeton University Press.
- Nakagawa, Shinichi, and Eduardo A. Santos. 2012. "Methodological Issues and Advances in Biological Meta-Analysis." *Evolutionary Ecology* 26(5):1253–74.
- Pullin, Andrew S., and Gavin B. Stewart. 2006. "Guidelines for Systematic Review in Conservation and Environmental Management." *Conservation Biology* 20(6):1647–56.
- Pullin, Andrew S., Teri M. Knight, David A. Stone, and Kevin Charman. 2004. "Do Conservation Managers Use Scientific Evidence to Support Their Decision-Making?" *Biological Conservation* 119(2):245–52.
- Pynegar, Edward L., James M. Gibbons, Nigel M. Asquith, and Julia P. Jones. 2021. "What Role Should Randomized Control Trials Play in Providing the Evidence Base for Conservation?" *Oryx* 55(2):235–44.
- Salafsky, Nick, Judith Boshoven, Zuzana Burivalova, Natalie S. Dubois, Andres Gomez, Arlyne Johnson, Aileen Lee, Richard Margoluis, John Morrison, Matthew Muir, Stephen C. Pratt, Andrew S. Pullin, Daniel Salzer, Annette Stewart, William J. Sutherland, and Claire F. Wordley. 2019. "Defining and Using Evidence in Conservation Practice." *Conservation Science and Practice* 1(5):e27.
- Shackelford, Gorm E., Philip A. Martin, Amelia S. Hood, Alex P. Christie, Elena Kulinskaya, and William J. Sutherland. 2021. "Dynamic Meta-Analysis: A Method of Using Global Evidence for Local Decision Making." *BMC Biology* 19(1):1–13.
- Stegenga, Jacob 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.
- Stegenga, Jacob 2018. *Medical Nihilism*. Oxford: Oxford University Press.
- Sutherland, William J., and C. F. Wordley. 2017. "Evidence Complacency Hampers Conservation." *Nature Ecology & Evolution* 1(9):1215–16.
- Vieland, Veronica J., and Hasok Chang. 2019. "No Evidence Amalgamation Without Evidence Measurement." *Synthese* 196(8):3139–61.

- von Wernsdorff, Melina, Martin Loef, Brunna Tuschen-Caffier, and Stefan Schmidt. 2021. "Effects of Open-Label Placebos in Clinical Trials: A Systematic Review and Meta-Analysis." *Scientific Reports* 11(1):1–14.
- Whittaker, Robert J. 2010. "Meta-Analyses and Mega-Mistakes: Calling Time on Meta-Analysis of the Species Richness–Productivity Relationship." *Ecology* 91(9):2522–33.
- Wüthrich, Nicolas, and Katie Steele. 2019. "The Problem of Evaluating Automated Large-Scale Evidence Aggregators." *Synthese* 196(8):3083–102.