This postprint has been accepted for publication in the journal *Science of the Total Environment*

Acting in the Face of Evidentiary Ambiguity, Bias, and Absence Arising from Systematic Reviews in Applied Environmental Science

Trina Rytwinski[a*], Steven J. Cooke[a], Jessica J. Taylor[a], Dominique G. Roche[a], Paul A. Smith[b,c], Greg W. Mitchell[b,c], Karen E. Smokorowski[d], Kent A. Prior[e], Joseph R. Bennett[a]

[a]Canadian Centre for Evidence-Based Conservation, Department of Biology and Institute of Environmental and Interdisciplinary Science, Carleton University, 1125 Colonel By Dr., Ottawa, ON, K1S 5B6, Canada

[b]Wildlife Research Division, Environment and Climate Change Canada, National Wildlife Research Centre, 1125 Colonel By Dr., Ottawa, ON, K1S 5B6, Canada

[c]Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

[d]Great Lakes Laboratory for Fisheries and Aquatic Sciences. Fisheries and Oceans Canada. 1219 Queen St. E., Sault Ste. Marie, ON, P6A 2E5, Canada

[e]Parks Canada, 30 Victoria (PC-03-C), 3-84, Gatineau, QC, J8X 0B3, Canada

[*]Correspondence: trytwinski@hotmail.com

Abstract: Evidence-based decision-making often depends on some form of a synthesis of previous findings. There is growing recognition that systematic reviews, which incorporate a critical appraisal of evidence, are the gold standard synthesis method in applied environmental science. Yet, on a daily basis, environmental practitioners and decision-makers are forced to act even if the evidence base to guide them is insufficient. For example, it is not uncommon for a systematic review to conclude that an evidence base is large but of low reliability. There are also instances where the evidence base is sparse (e.g., one or two empirical studies on a particular taxa or intervention), and no additional evidence arises from a systematic review. In some cases, the systematic review highlights considerable variability in the outcomes of primary studies, which in turn generates ambiguity (e.g., potentially context specific). When the environmental evidence base is ambiguous, biased, or lacking of new information, practitioners must still make management decisions. Waiting for new, higher validity research to be conducted is often unrealistic as many decisions are urgent. Here, we identify the circumstances that can lead to ambiguity, bias, and the absence of additional evidence arising from systematic reviews and provide practical guidance to resolve or handle these scenarios when encountered. Our perspective attempts to highlight that, with evidence synthesis, there may be a need to balance the spirit of evidence-based decision-making and the practical reality that management and conservation decisions and action is often time sensitive.

Keywords: Environmental evidence, Evidence-based decision-making, Evidence synthesis, Evidentiary uncertainty, Meta-analysis

1. Introduction

From transitioning to a low carbon future (Hanley et al., 2018), to restoring degraded habitats (Aronson and Alexander, 2013), or from bending the curve for biodiversity loss (Mace et al., 2018), to improving waste management in developing countries (Bartone and Bernstein, 1993), "good" decisions need to be made that benefit the environment and humanity. Some issues and decisions are local in scale (e.g., what to do at a given site in a particular region) while others are national or global [e.g., United Nations (UN), Sustainable Development Goals (SDGs), Conference of the Parties (COP), Intergovernmental Panel on Climate Change (IPCC), International Union for Conservation of Nature (IUCN)]. In contemporary civil society, we expect and even demand that environmental decisions are based on the best available evidence (Sutherland et al., 2004). The concept of evidence-based decision-making is intuitive to scientists, but there are many reasons why it can be difficult to achieve in practice (Head, 2010; Oliver et al., 2014; Head, 2016).

A fundamental tenet of evidence-based decision-making is that there is some form of evidence synthesis that collates evidence and identifies emergent patterns that guide decision-makers. There are many forms of evidence synthesis (see Bilotta et al., 2015; Haddaway et al., 2015; Pullin et al., 2016; Cook et al., 2017) but the gold standard, in many cases, is a systematic review that leads to a quantitative meta-analysis. Systematic reviews differ from traditional literature reviews in that they are repeatable, transparent, comprehensive (incorporating not just peer reviewed findings but also relevant grey literature), and attempt to minimize bias through a critical appraisal phase (see Table 1 for term definitions). Systematic reviews are particularly attractive to decision makers when evidence from different sources conflicts, especially if accompanied by a quantitative synthesis that can weight the conflicting evidence according to some measure of its reliability (e.g., the inverse of the effect size variance). Systematic reviews can also provide transparent and objective assessments where topics are controversial or high profile e.g., environmental effects of pollution from mines or microplastics (Haddaway and Pullin, 2014). For example, in health sciences, systematic reviews are widely embraced and serve as the foundation for modern public health actions and medical interventions (see Cochrane Collaboration; https://www.cochrane.org/; Lavis, 2009).

In an ideal world, environmental management decisions would be supported by a large amount of evidence derived from robust studies with consistent findings, strong effect sizes and a strong, universally applicable signal emerging from a quantitative meta-analysis. Yet, it is common for systematic reviews to conclude that the evidence base is too small to enable meta-analysis (either due to lack of studies or studies being excluded due to poor reporting – e.g., no variance or sample sizes provided), or that the evidence base is of low reliability (e.g., biased in various ways such as lacking controls, baseline data prior to intervention(s), inadequate sample sizes). For example, Cook et al. (2014) investigated the contribution of systematic reviews to environmental management and conservation decisions and found that, of the 43 they reviewed, the strict eligibility criteria for reviews and the limited quality of much of the available primary literature led to a median of only 12% of relevant studies being included in the meta-analysis. In turn, these types of constraints can lead to results that are not robust (i.e., only a narrative analysis is possible), not generalizable (e.g., a different measure used by primary studies or indirect response), and/or highly variable with respect to the meta-analytical results. Beyond these limitations, the time, funding, and technical expertise required to conduct systematic reviews can be considerable.

In many ways, the outcome(s) of a systematic review depend entirely on the scientific rigour of the available evidence base. Some forms of interventions and studies will never lack bias given inherent limitations, particularly with respect to identifying appropriate controls or replicates in natural systems. Indeed, demonstrating causal relationships between stressors and responses in environmental systems is challenging because of the natural variability in environmental responses and the difficulties associated with performing rigorous experiments [e.g., lack of before-impact data, poor control matching, flawed units of replication (i.e., pseudoreplication), an inability to randomize treatments, and the presence of uncontrolled confounding factors] (Beyers, 1998; Downes et al., 2002; Norris et al., 2005, 2012; Nichols et al., 2017). These issues weaken our ability to infer with confidence that any observed biological impairment is caused by the suspected environmental stressor, or that an ecological recovery resulted from the management intervention designed to mitigate the impairment (Downes et al., 2002). When a systematic review results in an evidence base that is ambiguous, biased, or that provides little or no additional evidence to help inform management decisions, authors tend to focus on recommendations to improve the quality of primary research, which is unsatisfying and unacceptable to most commissioners/funders. Here, we provide practical guidance for acting in the face of evidence from systematic reviews that is ambiguous, biased, and/or absent.

2. Practical guidance

We acknowledge that there are many considerations taken into account prior to deciding on the use of a systematic review in the decision-making process. Although these considerations are not our focus, we think it would be helpful to first outline them below. Then, for the situation where a systematic review is chosen as the appropriate tool, we provide guidance on what to do when the evidence base for the systematic review leads to ambiguity, bias, or a lack of additional information beyond what is already contained in the primary studies. This discussion piece was not intended to act as a recipe guide *per se* to resolve or handle these scenarios when encountered; rather, we outline potential options, providing a few specific examples where possible and appropriate. Our focus was not on directly tackling the issue of *how* to implement decisions when faced with evidence that is ambiguous, biased or when no additional evidence arises. Instead, we aimed to describe the circumstances that lead to these issues and offer some guidance to navigate the decision-making process.

The first step in evidence-based decision-making is identifying the need for evidence relating to a question of concern in policy or management. Often, questions stemming from discussions of evidence needs start out very broad, and occasionally are not well defined (Game et al., 2013; CEE, 2018). In some cases, the scale and scope of a problem may be such that it is obvious that local data will be most important for making decisions. Therefore, constructing a clear, carefully formulated question is essential. In this regard, there are a number of highly informative resources available for guidance (e.g., Gregory et al., 2012; Groves and Game 2015; Hammon et al., 2015; CEE, 2018).

Once the question has been properly formulated, the appropriate framework needs to be selected in a thoughtful way. To this end, there are also several useful tools for evidence-based decision making in environmental science. However, as noted by others (e.g., Pressey et al., 2013; Bower et al., 2018), deciding on which tool to use and how to implement it can be challenging for practitioners. Recently, four papers have provided guidance on how to address these challenges.

Schwartz et al. (2018) describe and contrast different planning and decision frameworks for systematic decision making. Bower et al. (2018) provide guidance on how to choose among three of the most common types of frameworks for solving environmental/conservation problems (i.e., structured decision making, systematic prioritization, and evidence synthesis), and how to identify less rigorous techniques when there are time or data availability constraints. Salafsky et al. (2019) provide a typology of the different kinds of evidence a project team requires to help make the various decisions needed to iteratively go through the decision-making process. Finally, Wright et al. (2020) provide potential actionable steps for bridging the gap between decision identification and action implementation (i.e., ‘decision-implementation gap’) as well as avenues for future development of decision frameworks.

If it is decided that the question of interest requires a synthesis of previous findings (based on guidance from sources such as those noted above), it is important to first check whether an evidence synthesis already exists, such as a systematic review or map (e.g., Collaboration for Environmental Evidence (CEE) syntheses library: https://www.environmentalevidence.org/completed-reviews), a subject-wide evidence synthesis [e.g., Sutherland et al. (2019); https://www.conservationevidence.com/], or a stand-alone meta-analysis. If there is an existing evidence synthesis, one can make use of a new online, freely available CEE evidence service known as CEEDER (http://www.environmentalevidence.org/ceeder; Konno et al., 2020). With this database, one can search evidence syntheses [commercially published reviews (available now) and grey literature reviews (forthcoming)] on a specific question of environmental policy or management relevance, and also obtain along with them an independent assessment of the reliability of each synthesis with respect to its use in decision-making. However, if there is no pre-existing evidence synthesis, or a more reliable or more up-to-date one is needed, then one must decide on the appropriate type of synthesis.

There are many approaches to evidence synthesis (described in Dicks et al., 2014; Bilotta et al., 2015; Haddaway et al., 2015; Pullin et al., 2016; Cook et al., 2017; Dicks et al., 2017; Sutherland and Wordley, 2018). Previous authors have provided guidance on which approach to use considering the type of question, policy context, desired outcomes of the synthesis (e.g., level of certainty required, level of transparency/repeatability required) and constraints on decision-makers (e.g., the available funding, level of technical expertise, deadlines) (see Pullin et al., 2016; Cook et al., 2017; Dicks et al., 2017). Available approaches to evidence synthesis can be viewed on an approximate continuum of very low rigor to very high rigor, and relatedly, limited usefulness to very useful with respect to their ability to inform management decisions. As mentioned previously, systematic reviews sit on that very high rigor end of the spectrum because they incorporate mechanisms to minimize bias in searching, study inclusion, critical appraisal, and meta-analytical sensitivity analyses, as well as increase transparency/reproducibility. However, systematic reviews require considerable resources. For instance, Haddaway and Westgate (2019) estimated that the average CEE systematic review takes 164 days at one full-time equivalent including vacations/holidays (but not weekends) and other regular disruptions (standard deviation=23 days). Note that this average estimate represents resource requirements in person days (i.e., the average number of days a project lead is working on the project) and not the total time it would take for a systematic review project to be completed (i.e., including time for journal assessment of the protocol and review), which – as identified by Haddaway and Westgate (2019) – is approximately 737 days (standard deviation=364 days). Therefore, if resource requirements go beyond the time (and/or budget) available to make a management decision, one may want to consider a more rapid method of synthesis; however, this decision may come at a cost of lower confidence in synthesis results (see Cook et al., 2017 for a helpful decision tree). As noted by Cook et al. (2017), “The challenge is to select an approach that maximises the efficiency, appropriateness and effectiveness of the resources used in the review process to deliver conclusions, with a sufficient level of certainty for the decision context”.

When a systematic review approach is deemed to be appropriate, there are still considerations that must be addressed. For instance, at this stage, it is sometimes unclear what the evidence base for the given topic actually resembles (i.e., Realistically how large is it? Is it generally reliable? How broad or narrow is its scope and/or scale?), which can impact the decision of whether a systematic review is in fact appropriate. If a systematic review proceeds, will there be sufficient, unbiased evidence to make conclusions on management outcomes or will it only be able to identify knowledge gaps and make recommendations on how to improve the validity of the evidence base? Based on the guidelines set out by the CEE, early stages of a systematic review involve a scoping exercise to develop the search string and test for search comprehensiveness (see Conducting a Search in CEE, 2018). However, this initial scoping exercise does not always provide a clear indication of how large or reliable the evidence base is - i.e., of the total number of studies found, how many studies are actually relevant to the management question (either directly and/or indirectly), nor does it provide insights as to the reliability of those primary studies - i.e., of those that are relevant, how many are unbiased? Therefore, if a systematic review is believed to be needed and if time/resources permit, we recommend doing a more rigorous scoping exercise to get an estimate of the size and reliability (internal and external study validity; refer to Table 1) of the evidence base. If the systematic review is commissioned by decision-makers, this would ideally be a separate contract, before deciding on what tool/framework is likely more appropriate to address the environmental management question. Here, the scoping exercise would first involve following the full CEE guidelines for article searches (see Conducting a Search in CEE, 2018). Then, using a subset of articles captured by the search, including commercially published AND grey literature sources, articles can be screened using pre-defined eligibility criteria (e.g., specific population or intervention of interest), to identify relevant sources of evidence. From here, an estimate of the inclusion rate can be made (i.e., of the number of articles in the subset, how many were deemed relevant), and used to predict how many articles from the full search results could be relevant to the review. Additionally, it is then possible to gauge the likely reliability of the full evidence base by estimating how many of those articles deemed relevant from the subset were found to be credible (e.g., overall high, medium, low study validity) In doing so, one can get an approximate estimate of the size and reliability of the evidence base, which can inform a decision as to the appropriateness of a systematic review. To our knowledge, no one has previously suggested nor attempted this form of a scoping exercise prior to conducting an evidence synthesis in environmental science (providing an example using hypothetical data is beyond the scope of this discussion piece). Assuming the evidence base is reliable, having an estimate of the size of the evidence base will help shape expectations around timelines and costs for carrying out the full systematic review.

Regardless of whether reliability of the evidence base is determined on the front-end with a scoping exercise as suggested above or on the back-end of an ongoing or completed systematic review, it is important for researchers to recognize five potential scenarios that will influence the strength of the conclusions that are drawn from the systematic review. We describe these scenarios directly below and paths forward when these scenarios are encountered:

i. If the evidence base is large, narrow in focus, and has relatively high reliability, the conclusions drawn from the systematic review exercise should not be limited by issues of evidentiary bias, or absence of additional evidence. However, there may be issues of evidentiary ambiguity; in this case, see *2.1 What to do when the evidence base is ambiguous?* below.

ii. If the evidence base is large, narrow in focus, and has mixed reliability, the conclusions drawn from the systematic review exercise may be limited by evidentiary ambiguity and/or bias; in this case, see *2.1 What to do when the evidence base is ambiguous?* AND *2.2 What to do when the evidence base is biased?* below.

iii. If the evidence base is large, narrow in focus, and has generally low reliability, the conclusions drawn from the systematic review exercise will be primarily limited by evidentiary bias and a different framework/tool should be considered instead of a systematic review (see Cook et al., 2017; Bower et al., 2018) but also see *2.2 What to do when the evidence base is biased?* below.

iv. If the evidence base is deemed large but broad in scope/scale/outcome types, the conclusions drawn from the systematic review may be limited by evidentiary ambiguity. If this limitation is identified on the front-end of a systematic review, one can:

   a. Consider a systematic map as a starting point to generate a database and identify knowledge gaps (i.e., primary research needs) and clusters (i.e., areas for future systematic reviews) (see CEE, 2018).

i. The systematic map can also be combined with Multiple Expert Consultation + Delphi method to analyze evidence over a broad area relatively quickly (Dicks et al., 2017).

b. Consider narrowing the scope (e.g., select a clear knowledge cluster from scoping effort and focus the systematic review on that topic).

If this limitation is identified during the systematic review process (i.e., back-end), see *2.1 What to do when the evidence base is ambiguous?* below.

v. If the evidence base is sparse, the conclusions drawn from the systematic review will be limited by the absence of additional information. If this limitation is identified on the front-end of a systematic review (i.e., during the scoping exercise), one can:

a. Consider broadening the scope of the review to capture more evidence (e.g., multiple forms of interventions and outcomes).

b. Consider broadening the review search to include openly accessible datasets to make use of additional data from non-target studies that have attained relevant information to address different research questions (see Culina et al., 2018).

c. Revisit other frameworks/tools (i.e., Multiple Expert Consultation + Delphi method) (see Dicks et al., 2017 and Bower et al., 2018).

d. Proceed with the systematic review but acknowledge and communicate clearly with practitioners the limitations of the current evidence base to manage expectations (i.e., inform them that meta-analysis will not be possible and therefore, the synthesis will take the form of a narrative synthesis); in this case, see *2.3 What to do when there is no additional evidence?* below.

If this limitation is identified during the systematic review process (i.e., back-end), the conclusions drawn from the systematic review will also be limited by the absence of additional information; therefore, see *2.3 What to do when there is no additional evidence?* below.

2.1 What to do when the evidence base is ambiguous?

In science, the term uncertainty is often treated as a single concept that simply represents the absence of precise information (Molden and Higgins, 2004). However, important distinctions have been made between different varieties of uncertainty. One such variety, ambiguous uncertainty, is a term that is commonly used but is not easily defined. This is because scientists sometimes use common words to mean different things but also different varieties of uncertainty are not mutually exclusive. For instance, Molden and Higgins (2004) describe ambiguous uncertainty as an abundance of conflicting information regarding a possible decision. Whereas Smith and Stern (2011) describe ambiguity as being related to outcomes for which probability statements cannot be provided (i.e., arising when there are impacts whose uncertainty one cannot quantify via probabilities; also known as Knightian uncertainty). They also acknowledge that ambiguity sometimes reflects uncertainty in an estimated probability (i.e., imprecision uncertainty). A key difference between these types of uncertainty is that the impacts of imprecision uncertainty on decisions can be more easily explored via sensitivity analysis (e.g., Tulloch et al., 2013). An important point noted by Smith and Stern (2011) is that, while science aims to reduce ambiguity and quantify imprecision, there is not always a clear distinction between the two. What matters in the context of evidence synthesis is that ambiguity arising from the evidence base and results of a systematic review, whether due to quantifiable or unquantifiable uncertainty, can translate into ambiguity in terms of the appropriate decision or management response (Faucheux and Froger, 1995).

To address ambiguity arising from systematic reviews, we first encourage that researchers and decision-makers embrace and accept the fact that decisions are almost never final, particularly at large spatial scales (species extinctions are an obvious exception to this). Decisions are often revisited, changed, or cancelled based on the accumulating evidence or its interpretation in different socio-political frameworks. Furthermore, it may be necessary to acknowledge that some decisions cause synergistic and some antagonistic responses given the complexity of biological and human responses to change (e.g., Folt et al., 1999; Côté et al., 2016). To that end, it is worthwhile establishing dynamic processes that re-evaluate evidence as new evidence becomes available or contexts change (Gonzalez, 2005). We also encourage decision makers to consider an adaptive management approach that incorporates or studies the outcome of a particular management decision, which can be used to update the evidence base itself. Evidence synthesis is best achieved when new evidence is incorporated into the evidence base as it becomes available. Currently, there is no established CEE framework for this process; however, it is common in the healthcare field and guidance has been developed in that realm (Moher and Tsertsvadze, 2006; Garner et al., 2016; also see 3 Final remarks for suggestions for future work below). This has become particularly salient during the COVID-19 pandemic where vast amounts of new knowledge are being generated rapidly (Tricco et al., 2020), with lessons emerging that are relevant to environmental evidence synthesis (Kadykalo et al., 2021).

In addition to incorporating new evidence as it becomes available, there are other ways of exploring or reducing ambiguity in systematic reviews. For instance, it is common for an evidence base to be mixed (i.e., having a blend of positive and negative evidence). Although meta-analysis serves to combine these effects and obtain an estimate of the mean overall effect, the variability around this mean can be high. Relationships between potential sources of heterogeneity and effect size estimates can and should be explored as part of the meta-analysis. However, these analyses are generally easier and more appropriate to undertake when there is a large evidence base to reduce Type I (false positives) and II (false negatives) errors (CEE, 2018). Furthermore, in some situations, the evidence base may contain broadly different outcomes, management interventions, and/or taxa which could make studies inadequately comparable when attempting to pool results in a meta-analysis. In these situations, heterogeneity and the potential for ambiguity can be reduced by partitioning studies into more appropriately comparable subgroups (e.g., different outcomes) and conducting distinct meta-analyses. Potential differences in study characteristics can then be explored within these separate subgroups via the inclusion of moderators.

Ambiguity may also arise as a result of the decisions made regarding how the meta-analysis was conducted. As noted by Haddaway and Rytwinski (2018), each step in conducting a meta-analysis requires decisions that have both scientific and statistical implications. When meta-analyzing evidence, researchers are often faced with a number of decisions (e.g., choice of effect size measure, variance calculations, model building, analysis software) and sometimes must choose between equally valid approaches. Some of these meta-analytical decisions are subjective, which can have implications on analysis results and lead to ambiguity. Therefore, it is critical that researchers comprehensively and transparently report their methodology (i.e., what decisions were made and why), and for journal editors and evidence synthesis coordinating bodies (e.g., CEE) to ensure that quantitative synthesis methods are adequately reported and justified in published systematic reviews. Furthermore, Haddaway and Rytwinski (2018) advocate that, when possible, reviewers should attempt analyses in multiple ways if two or more equally valid approaches are possible to see how results compare, presenting results within a range of uncertainty when results conflict or differ. We acknowledge that regardless of whether exploring ambiguity with moderator analyses or the meta-analytical choices made to summarize the evidence, ambiguity may still remain. Therefore, we reiterate the importance of continuing to incorporate new evidence as it becomes available, including new original research (e.g., on additional sites or interventions) directed by the outcomes of ambiguous systematic reviews. To do so requires that decision-maker and decision-making bodies are equipped to deal with dynamic processes and willing to embrace the concept that most decisions are not final.

2.2 What to do when the evidence base is biased?

CEE (2018) defines bias (i.e., internal validity) as "a systematic deviation in study results from their true value, i.e., either an underestimation or overestimation of the true value". Unlike statistical uncertainty due to random error (present in all studies), bias as a result of a systematic error cannot be overcome by increasing sample sizes in a given study or by combining study results in a meta-analysis. If bias is present in primary studies, their results will be incorrect. Subsequently, if a systematic review is based on incorrect evidence, the results of the meta-analysis will also be incorrect, resulting in misleading conclusions (Boutron et al., 2019). For example, a misleading conclusion could stem from a systematic review where a precise but wrong answer is made. Directly measuring bias within primary studies is challenging. Instead, for systematic reviews, an indirect approach is used to infer the "risk of bias" by examining aspects of research conduct (i.e., study design and methods) to determine whether studies used adequate methodology to protect against bias (i.e., often referred to as critical appraisal of study validity) (Higgins et al., 2011; CEE, 2018). To do so, researchers generally use review-specific assessment criteria for appraising the interval validity – developed at the protocol stage and ideally in consultation with topic experts and relevant stakeholders – categorizing studies, for example, as having overall high, medium or low validity.

When the evidence base of a systematic review is assessed to have mixed reliability (i.e., the evidence base is made up of both higher and lower risk of bias studies), it is important to understand the potential impact of this bias on review results (Boutron et al., 2019). To do this, reviewers should test the influence of including studies of higher risk of bias on the review results by means of sensitivity analysis. For example, if the evidence base allows for meta-analysis, one could stratify studies according to the overall risk of bias to produce and compare multiple effect estimates from models that include, for example, all studies, studies at lower risk of bias only, and studies at higher risk of bias only. Sensitivity analysis could be used to make decisions as to whether (1) the meta-analysis should be restricted to studies at low risk of bias when it seems clear from the model comparisons that the conclusions are likely impacted by the inclusion of studies at high risk of bias, or (2) multiple effect estimates for different risk of bias stratifications should be presented. The limitation associated with the former approach is the potential for the loss of precision when excluding high risk of bias studies from the analysis (i.e., not making full use of the evidence base). If there are relatively few studies with high risk of bias but these studies have a clear impact on the mean effect estimate, excluding these few studies would seem like a valid trade-off to achieve a result that is unbiased but potentially less precise. However, if there are only a few studies at low risk of bias, excluding all studies with high risk of bias may produce a result that is unbiased but imprecise, which may not be a valid trade-off. In the latter approach (i.e., 2), while the impact of bias on review results is presented, the limitation is reporting multiple effect estimates for a given outcome which may be confusing, especially for decision-makers if they are looking for a single result. However, one option to address this is to present a stratified (ordered) forest plot displaying all the information transparently [e.g., a forest plot displaying the effect size estimates of each study included in the meta-analysis stratified by the overall risk of bias judgment; see Fig. 1]. When the majority of analyses trend towards the same conclusion, even if overall effect size estimates and significance vary, this can provide greater support of decisions than a single overall quantitative estimate from limited low risk of bias studies.

When the evidence base consists largely (or entirely) of low reliability studies, most often formal meta-analytical procedures are not possible. In such cases, the evidence base is usually only discussed narratively (i.e., tabulation and/or visualization). Although we have made a case that systematic reviews conducted in accordance with international standards (CEE, 2018) are the gold standard for evidence synthesis, the reality is that the existing literature base for some environmental science topics are such that they will never be free of bias. Recognizing this, we advocate, as others have done (e.g., Doerr et al., 2015), that researchers should strive to go beyond a simple narrative synthesis and attempt some form of an analysis of the primary studies even if formal rigorous meta-analytical methods are not possible [e.g., a sign test, meta-analyses of p-values (see Borenstein et al., 2009); meta-analyses of single arm proportions for non-comparator studies (see Lipsey and Wilson, 2001), meta-analysis using an alternate effect size metric such as percent change in intervention effectiveness (e.g., Rytwinski et al., 2019; see Box 1)]. A potentially less rigorous analysis with clear caveats and a discussion of the resultant implications on the review findings will provide better, more usable information than no analysis at all. However, if this path is chosen, it needs to be done acknowledging uncertainty and with future efforts focused on improving and expanding the evidence base.

It is also important to recognize the fundamental importance of evaluating and synthesizing evidence irrespective of whether a systematic review results in a quantitative analysis. Systematic reviews generate a curated database of nearly all relevant evidence sources, which is a highly valuable resource (e.g., Conservation Evidence; https://www.conservationevidence.com/). Even when the evidence base as a whole is deemed to be of low reliability, any and all evidence about the threat or role of interventions could help to tip the scales in the direction of a good decision. Indeed, others have also advocated for considering all forms of evidence appropriately when informing policy and practice (e.g., Sutherland and Wordley, 2018; Salafsky et al., 2019). The benefit of a systematic review, as outlined above, is that study validity is assessed for each study such that if it is deemed of low reliability, one is told why that is the case. In that sense, it is very much a "user be warned" message. All this is to say that the database of existing studies when combined with detailed information on study validity can play an important role in informing decisions.

2.3 What to do when there is no additional evidence?

In rare cases, systematic reviews may produce no additional evidence that can be used to address a problem. Quoting the co-founder of CEE "evidence synthesis can't make sense out of nothing" (A. Pullin, pers. comm.). Given the resource requirements for systematic reviews, this scenario can be disappointing for both commissioners/funders and authors of the systematic review. Ideally, it can be avoided by careful problem formulation (cf. Gregory et al., 2012; Bower et al.,

2018) to ensure the problem to be explored in the systematic review is clear and answerable. Scoping exercises (outlined above) in the early stages of a systematic review also offer an opportunity to identify questions for which no useful evidence is likely to be found. However, the complexities of environmental decisions, whereby many components interact on different spatiotemporal scales, mean that the problem of no additional relevant evidence may still arise because little or no research has been conducted at a relevant spatial, temporal or taxonomic resolution.

Such a situation may offer key opportunities since no additional evidence may still amount to important additional information. First, it may indicate that the problem formulation process should be re-visited to further refine the question toward something that is both relevant and answerable. This outcome may not have been predictable at the onset of the project. Second, finding no additional evidence may provide a strong mandate to act on current information, despite uncertainties. Although environmental managers tend to be risk-averse (Tulloch et al., 2015), sometimes rapid action is crucial, despite high uncertainty (Martin et al., 2012). Finally, no additional evidence may indicate that original field or lab research is necessary to address the problem. In this case, additional tools can help focus the research on the spatiotemporal scales that will inform management decisions. Value of information theory (Raiffa, 1968) is a powerful tool to help focus research so that it is as informative for decisions as possible (Runge et al., 2011; Bennett et al., 2018). For example, Raymond et al. (2020) found that optimally locating surveys for threatened plant species using value of information theory would allow managers to protect more habitats with their limited resources. They also found that for many situations, acting on current information was more efficient than gathering more evidence. Similarly, Maxwell et al. (2015) used value of information theory to show that new data would have negligible impact on management decisions for halting koala population declines. Bayesian belief networks, which incorporate uncertainty into interactive models of systems, can also be used to identify key areas of uncertainty that can most influence decisions (e.g., McCann et al., 2006; Howes et al., 2010). Bayesian belief networks can also incorporate many different forms of data. For example, Smith et al. (2007) predicted suitable habitat for the Julia Creek Dunnart (*Sminthopsis douglasi*), an endangered marsupial, using a Bayesian belief network that incorporated expert elicitation regarding habitat use (for which there was little detailed information), remotely-sensed proxies for key environmental variables, and confirmatory data from fieldwork.

3. Final remarks

We acknowledge there are future guidance needs with respect to improving evidence syntheses in environmental science. For instance, to date, there have been relatively few rigorous methods proposed or developed to include different ways of knowing (e.g., stakeholder, practitioner or Indigenous knowledge) in formal evidence-based decision-making processes. Fortunately, there are many groups working in this space (see Berkes, 2009; Phillipson et al., 2012; Haddaway et al., 2019) and there are huge dividends to be realized should we be able to figure out how to do so (Hulme, 2010). Yet, it is important to acknowledge that decisions based on “evidence” often fail to recognize that most knowledge holders do not present their work in either the grey or peer-reviewed literature. This does not mean that those sources of knowledge are any less valid. Indeed, in many cases they are the only or best source of knowledge. What is lacking from current evidence synthesis approaches are mechanisms to formally bridge those ways of knowing with other traditional western science methods while simultaneously accounting for bias in all ways of knowing. The “two-eyed seeing” approach, which briefly, encourages that we learn to see from one eye with the best in the Indigenous ways of knowing, and from the other eye with the best in the Western ways of knowing, and that we learn to use both these eyes together, for the benefit of all, is one of the first practical approaches for doing so (Bartlett et al., 2012). However, it has yet to be fully embraced or extended to include knowledge keepers or holders beyond Indigenous Peoples.

Furthermore, there are currently no guidelines available to provide a rigorous, transparent, and unbiased synthesis of the literature to address more urgent environmental management/policy questions (i.e., 1-2 months). The shortest currently available well-defined methods take two or more months (i.e., Quick Scoping Reviews or Rapid Evidence Assessments) and lie between regular literature reviews and systematic reviews in terms of rigour of assessment (Collins et al., 2015). Therefore, guidelines for rapid synthesis approaches need further attention. In the meantime, if policy makers need to rely on less rigorous methods of evidence synthesis (e.g., regular literature reviews, vote counting), it is essential that they are accompanied with clear caveats. Additionally, horizon scanning could be used so that evidence needs are anticipated in advance (Sutherland et al., 2020).

Further exploration into incorporating frameworks endorsed by other fields should also be considered. For example, health care sciences have adopted the GRADE approach (Grading of Recommendations, Assessment, Development and Evaluation; https://www.gradeworkinggroup.org/) to help move from the results of the systematic review to making conclusions and presenting the evidence to decision makers via summaries of evidence. Here, GRADE is used to rate the body of evidence at the outcome level rather than the study level, to provide an overall GRADE certainty rating (i.e., high, moderate, low and/or very low) to evaluate the strength of recommendations in order to assist decision makers. GRADE provides a reproducible and transparent framework for grading the certainty of evidence and strength of recommendations for medical science; how this system could be adapted to the environmental science realm deserves further consideration.

Furthermore, as touched on above in the section on *What to do when the evidence base is ambiguous?*, developing approaches to updating and incorporating new evidence into the evidence base as it becomes available is important to ensure systematic review are not at risk of inaccuracy (Shojania et al. 2007). One novel approach stemming from the healthcare field, that goes beyond simply updating the evidence base, is the concept of a living systematic review (Elliot et al., 2014). Elliot et al. (2017) describe a living systematic review in practice as the “continual surveillance for new research evidence through ongoing or frequent searches and the inclusion of relevant new information into the review in a timely manner so that the findings of the systematic review remain current”. In contrast to standard review updating, living systematic reviews include an explicit and *a priori* commitment to keeping the systematic review as current as possible with a predetermined frequency of search and review (e.g., most current living systematic review pilot projects aim to search most sources at least monthly and make the results of these searches visible to end users within another month) (Elliot et al., 2017). Living guideline recommendations, as well as guidance on statistical methods for updating meta-analyses have been developed for the health care realm (see Elliot et al., 2017; Simmonds et al., 2017). Therefore, how this approach could work and be modified in the environmental field deserves further consideration, especially where (1) evidence for particular topics are emerging rapidly, (2) current evidence is ambiguous, and (3) new search may change policy or practice. Furthermore, how this approach could be maintained and supported long-term with respect to the required continual application of (modest) resources also deserves further attention.

In conclusion, in environmental management and conservation, there are many cases where the evidence base is of sufficiently high validity and size to accommodate systematic reviews and where conclusions from systematic reviews have been instrumental in informing policy and practice (see Haddaway and Pullin, 2014). However, the reality is that there are circumstances when the evidence base is simply vague, limited in size/scope, and/or affected by (unavoidable) biases. Here, we have provided practical guidance for how to resolve or handle circumstances that can lead to ambiguity, bias, and the absence of additional evidence arising from systematic reviews (summarized in Box 2). We hope this advice will reinforce the idea that systematic reviews are part of a suite of decision tools, which can inform each other (cf. Bower et al., 2018), and that ambiguous, biased or no additional evidence arising from a systematic review can still be an important outcome for decisions. Our perspective attempts to highlight that, in some situations, there is a need of a balance between the spirit of evidence-based decision making (i.e., using only the most rigorous studies in evidence synthesis to inform decisions) and the practical reality that rapid action is often crucial for environmental management and conservation (i.e., using what evidence is available now while identifying and/or minimizing ambiguity, bias, and the absence of additional evidence arising from systematic reviews). This perspective does not however provide practical advice for decision makers on *how* to implement a decision when faced with these circumstances, as this was beyond the scope of our discussion. We acknowledge that filling such a gap with practical guidance (e.g., establishing frameworks, standardized processes) would be vital for those tasked with making environmental management and conservation decisions, and as such deserves immediate consideration.

CRediT authorship contribution statement

Trina Rytwinski: Conceptualization, Validation, Visualization, Writing - original draft, Writing - review & editing. **Steven J. Cooke:** Conceptualization, Validation, Visualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Jessica J. Taylor:** Conceptualization, Validation, Writing - review & editing. **Dominique G. Roche:** Validation, Writing - original draft, Writing - review & editing. **Paul A. Smith:** Validation, Writing - review & editing. **Greg W. Mitchell:** Validation, Writing - review & editing. **Karen E. Smokorowski:** Conceptualization, Validation, Writing - review & editing. **Kent A. Prior:** Validation, Writing - review & editing. **Joseph R. Bennett:** Conceptualization, Validation, Visualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

This work was supported by Environment and Climate Change Canada [grant number GCXE19S058]. Additional support was provided by Carleton University and the Natural Sciences and Engineering Research Council of Canada. We also thank the reviewers for their constructive comments.

References

Aronson, J., Alexander, S. 2013. Ecosystem restoration is now a global priority: time to roll up our sleeves. Restoration Ecology 21(3): 293-296.

Bartlett, C., Marshall, M., Marshall, A. 2012. Two-eyed seeing and other lessons learned within a co-learning journey of bringing together indigenous and mainstream knowledges and ways of knowing. Journal of Environmental Studies and Sciences 2(4): 331-340.

Bartone, C.R., Bernstein, J.D. 1993. Improving municipal solid waste management in third world countries. Resources, Conservation and Recycling 8(1-2): 43-54.

Bennett, J.R., Maxwell, S.L., Martin, A.E., Chadès, I., Fahrig, L., Gilbert, B. 2018. When to monitor and when to act: value of information theory for multiple management units and limited budgets. Journal of Applied Ecology 55(5): 2102-2113.

Berkes, F. 2009. Indigenous ways of knowing and the study of environmental change. Journal of the Royal Society of New Zealand 39(4): 151-156. http://dx.doi.org/10.1080/03014220909510568

Beyers, D.W. 1998. Causal inference in environmental impact studies. Journal of the North American Benthological Society 17: 367–373.

Bilotta, G.S., Milner, A.M., Boyd, I.L. 2015. How to increase the potential policy impact of environmental science research. Environmental Science Europe 27: 9. doi: 10.1186/s12302-015-0041-x.

Borenstein, M., Hedges, L.V., Higgins, J.P.T, Rothstein, HR. 2009. Meta-analysis methods based on direction and *p*-values. In: Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., editors. Introduction to meta-analysis. Hoboken, Wiley, p. 325–330.

Boutron I., Page, M.J., Higgins, J.P.T., Altman, D.G., Lundh, A., Hróbjartsson, A. 2019. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). http://www.training.cochrane.org/handbook. Accessed 31 May 2020.

Bower, S.D., Brownscombe, J.W., Birnie-Gauvin, K., Ford, M.I., Moraga, A.D., Pusiak, R.J.P., Turenne, E.D., Zolderdo, A.J., Cooke, S.J., Bennett, J.R. 2018. Making tough choices: picking the appropriate conservation decision-making tool. Conservation Letters 11: e1241. https://doi.org/10.1111/conl.12418

[CEE] Collaboration for Environmental Evidence. 2018. In: Pullin, A.S., Frampton, G.K., Livoreil, B., Petrokofsky, G., editors. Guidelines and standards for evidence synthesis in environmental management. Version 5. http://www.environmentalevidence.org/information-for-authors. Accessed 31 May 2020.

Collins, A., Miller, J., Coughlin, D., Kirk, S. 2015. The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide. Joint Water Evidence Group. https://www.gov.uk/government/publications/the-production-of-quick-scoping-reviews-and-rapid-evidence-assessments. Accessed 31 May 2020.

Cook, C.N., Possingham, H.P., Fuller, R.A. 2014. Contribution of systematic reviews to management decisions. Conservation Biology 27(5): 902-915.

Cook, C.N., Nichols, S.J., Webb, J.A., Fuller, R.A., Richards, R.M. 2017. Simplifying the selection of evidence synthesis methods to inform environmental decisions: a guide for decision makers and scientists. Biological Conservation 213: 135–145.

Côté, I.M., Darling, E.S., Brown, C.J. 2016. Interactions among ecosystem stressors and their importance in conservation. Proceedings of the Royal Society B: Biological Sciences 283(1824): 20152592. https://doi.org/10.1098/rspb.2015.2592.

Culina, A., Crowther, T.W., Ramakers, J.J.C., Gienapp, P., Visser, M.E. 2018. How to do meta-analysis of open datasets. Nature Ecology & Evolution 2: 1053-1056.

DFO. 2020. Science Advice on the Effectiveness of Spawning Habitat Creation for Substrate Spawning Temperate Fish. DFO Canadian Science Advisory Secretariat Science Advisory Report 2020/013. https://waves-vagues.dfo-mpo.gc.ca/Library/40876160.pdf. Accessed 9 December 2020.

Dicks, L.V., Haddaway, N., Hernández-Morcillo, M., Mattsson, B., Randall, N., Failler, P., Ferretti, J., Livoreil, B., Saarikoski, H., Santamaria, L., Rodela, R., Velizarova, E., Wittmer, H. 2017. Knowledge synthesis for environmental decisions: an evaluation of existing methods, and guidance for their selection, use and development – a report from the EKLIPSE project. https://www.eklipse-mechanism.eu/apps/Eklipse_data/website/EKLIPSE_D3-1-Report_FINAL_WithCovers_V6.pdf. Accessed 20 April 2020.

Dicks, L.V., Walsh, J., Sutherland, J.W. 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. Trends in Ecology and Evolution 29: 607-613.

Doerr, E.D., Dorrough, J., Davies, M.J., Doerr, V.A.J., McIntyre, S. 2015. Maximizing the value of systematic reviews in ecology when data or resources are limited. Austral Ecology 40: 1-11.

Downes, B.J., Barmuta, L.A., Fairweather, B.G., Faith, D.P., Keough, M.J., Lake, P.S., Mapstone, B.D., Quinn, Q.P. 2002. Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, Cambridge, UK.

Elliott, J.H., Turner, T., Clavisi, O., Thomas, J., Higgins, J.P.T., Mavergames, C., Gruen, R.L. 2014. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Medicine 11(2): e1001603. https://doi.org/10.1371/journal.pmed.1001603.

Elliott, J.H., Synnot, A., Turner, T., Simmonds, M., Akl E.A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton J., Tovey, D., Shemilt, I., Thomas, J., Living Systematic Review Network. 2017. Living systematic review: 1. Introduction-the why, what, when, and how. Journal of Clinical Epidemiology 91: 23-30. doi:10.1016/j.jclinepi.2017.08.010.

Faucheux, S., Froger, G. 1995. Decision-making under environmental uncertainty. Ecological Economics 15(1): 29-42.

Folt, C. L., Chen, C.Y., Moore, M.V., Burnaford, J. 1999. Synergism and antagonism among multiple stressors. Limnology and Oceanography 44(3part2): 864-877. https://doi.org/10.4319/lo.1999.44.3_part_2.0864.

Game, E.T., Kareiva, P. Possingham, H.P. 2013. Six common mistakes in conservation priority setting. Conservation Biology 27(3): 480-485.

Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Akl, E.A., Beyene, J., Chang, S., Churchhill, R., Dearness, K., Guyatt, G., Lefebvre, C., Liles, B., Marshall, R., Martínez García, L., Mavergames, C., Nasser, M., Qaseem, A., Sampson, M., Soares-Weiser, K., Takwoingi, Y., Thabane, L., Trivella, M., Tugwell, P., Welsh, E., Wilson, E.C., Schünemann, H.J. 2016. When and how to update systematic reviews: consensus and checklist. Research Methods and Reporting 354: i3507. http://dx.doi.org/10.1136/bmj.i3507

Gonzalez, C. 2005. Decision support for real-time, dynamic decision-making tasks. Organizational Behavior and Human Decision Processes 96(2): 142-154.

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T. Ohlson, D. (Eds.). 2012. Structured decision making: a practical guide to environmental management choices. Wiley-Blackwell, Oxford, UK.

Groves, C.R. Game, E.T. 2015. Conservation planning: informed decisions for a healthier planet. Roberts & Company, Greenwood Village, CO.

Haddaway, N.R., Pullin, A.S. 2014. The policy role of systematic reviews: past, present and future. Springer Science Reviews 2: 179-183.

Haddaway, N.R., Woodcock, P., Macura, B., Collins, A. 2015. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology 29(6): 1596-1605. doi: 10.1111/cobi.12541.

Haddaway, N.R., Bernes, C., Jonsson, B.G., Hedlund, K. 2016. The benefits of systematic mapping to evidence-based environmental management. Ambio 45(5): 613-620. doi: 10.1007/s13280-016-0773-x.

Haddaway, N.R., Rytwinski, T. 2018. Meta-analysis in not an exact science: call for guidance on quantitative synthesis decisions. Environment International 114: 357-359. https://doi.org/10.1016/j.envint.2018.02.018.

Haddaway, N.R., S.J. Cooke, P. Lesser, B. Macura, A.E. Nilsson, J.J. Taylor and K. Raito. 2019. Evidence of the impacts of metal mining and the effectiveness of mining mitigation measures on social–ecological systems in Arctic and boreal regions: a systematic map protocol. Environmental Evidence. 8:9. https://doi.org/10.1186/s13750-019-0152-8.

Haddaway, N.R., Westgate, M.J. 2019. Predicting the time needed for environmental systematic reviews and systematic maps. Conservation Biology 33: 434-443.

Hammond, J.R., Keeney, R.L. Raiffa, H. 2015. Smart choices: a practical guide to making better decisions. Harvard Business Review Press, Boston, MA.

Hanley, E.S., Deane, J.P., Gallachóir, B.Ó. 2018. The role of hydrogen in low carbon energy futures–A review of existing perspectives. Renewable and Sustainable Energy Reviews 82: 3027-3045.

Head, B.W. 2010. Reconsidering evidence-based policy: key issues and challenges. Policy and Society 29: 77-94.

Head, B.W. 2016. Toward more "evidence-informed" policy making? Public Administration Review 76(3): 472-484.

Higgins, J.P.T., Altman, D., Gøtzsche, P., Jüni, P., Mother, D., Oxman, A., Savović J., Schulz, K.F., Weeks, L., Sterne, J.A.C., Cochrane Bias Methods Group, Cochrane Statistical Methods Group. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343: d5928 doi: https://doi.org/10.1136/bmj.d5928.

Howes, A.L., Maron, M. Mcalpine, C.A. 2010. Bayesian networks and adaptive management of wildlife habitat. Conservation Biology 24(4): 974-983.

Hulme, M. 2010. Problems with making and governing global kinds of knowledge. Global Environmental Change 20(4): 558-564.

Jacaban, E., Rytwinski, T., Taylor, J.J., Young, N., Nguyen, V.M., Cooke, S.J. *In review*. Perspectives of environmental systematic review authors on the impact of their synthesis on policy and practice. Environmental Management.

James, K.L., Randall, N.P., Haddaway, N.R. 2016. A methodology for systematic mapping in environmental sciences. Environmental Evidence 5: 7. https://doi.org/10.1186/s13750-016-0059-6 [accessed 24 June 2020].

Kadykalo, A.N., N.R. Haddaway, T. Rytwinski and S.J. Cooke. 2021. Principles for generating accessible and useable COVID-19 environmental science and a fit-for-purpose evidence base. Ecological Solutions and Evidence 2: e12041. doi: 10.1002/2688-8319.12041

Konno, K., Cheng, S.H., Eales, J., Frampton, G., Kohl, C., Livoreil, B., Macura, B., O'Leary, B.C., Randall, N.P., Taylor, J.J., Woodcock, P., Pullin, A.S. 2020. The CEEDER database of evidence reviews: an open-access evidence service for researchers and decision-makers. Environmental Science and Policy 114: 256-262.

Lavis, J.N. 2009. How can we support the use of systematic reviews in policymaking? PLoS Medicine 6(11): e1000141. doi: 10.1371/journal.pmed.1000141

Lipsey, M.W., Wilson, D.B. 2001. Practical meta-analysis. SAGE Publications, Thousand Oaks.

Mace, G.M., Barrett, M., Burgess, N.D., Cornell, S.E., Freeman, R., Grooten, M., Purvis, A. 2018. Aiming higher to bend the curve of biodiversity loss. Nature Sustainability 1(9): 448-451.

Margules, C.R., Pressey, R.L. 2000. Systematic conservation planning. Nature 405(6783): 243-253.

Martin, T.G., Nally, S., Burbidge, A.A., Arnall, S., Garnett, S.T., Hayward, M.W., Lumsden, L.F., Menkhorst, P., McDonald-Madden, E. Possingham, H.P. 2012. Acting fast helps avoid extinction. Conservation Letters 5(4): 274-280.

Maxwell, S.L., Rhodes, J.R., Runge, M.C., Possingham, H.P., Ng, C.F., Mcdonald-Madden, E. 2015. How much is new information worth? Evaluating the financial benefit of resolving management uncertainty. Journal of Applied Ecology 52:12–20.

McCann, R.K., Marcot, B.G. Ellis, R. 2006. Bayesian belief networks: applications in ecology and natural resource management. Canadian Journal of Forest Research 36(12): 3053-3062.

Moher, D., Tsertsvadze, A. 2006. Systematic reviews: when is an update an update?. The Lancet 367(9514): 881-883.

Molden, D.C., Higgins, E.T. 2004. Categorization under uncertainty: resolving vagueness and ambiguity with eager versus vigilant strategies. Social Cognition 22(2): 248-277.

Mukherjee, N., Hugé, J., Sutherland, W.J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., Koedam, N. 2015. The Delphi technique in ecology and biological conservation: applications and guidelines. Methods in Ecology and Evolution 6: 1097-1109.

Nichols, S.J., Peat, M., Webb, J.A. 2017. Challenges for evidence-based environmental management: what is acceptable and sufficient evidence of causation? Freshwater Science 36(1): 240-249.

Norris, R.H., Liston, P., Mugodo, J., Nichols, S., Quinn, G.P., Cottingham, P., Metzeling, L., Perriss, S., Robinson, D., Tiller, D., Wilson, G. 2005. Multiple lines and levels of evidence for detecting ecological responses to management intervention. In: Rutherfurd, I.D., Wiszniewski, I., Askey-Doran, M.J., Glazik, R. (editors). Proceedings of the 4th Australian Stream Management Conference: Linking Rivers to Landscapes, Launceston, Tasmania. Department of Primary Industries, Water and Environment, Hobart, Tasmania. p. 456-463.

Norris, R. H., Webb, J.A., Nichols, S.J., Stewardson, M.J., Harrison, E.T. 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31: 5–21.

Oliver, K., Lorenc, T., Innvær, S. 2014. New directions in evidence-based policy research: a critical analysis of the literature. Health research policy and systems 12(1): 34. doi:10.1186/1478-4505-12-34.

Phillipson, J., Lowe, P., Proctor, A., Ruto, E. 2012. Stakeholder engagement and knowledge exchange in environmental research. Journal of environmental management 95(1): 56-65.

Pressey, R.L., Mills, M., Weeks, R., Day, J.C. 2013. The plan of the day: managing the dynamic transition from regional conservation designs to local conservation actions. Biological Conservation 166: 155-169.

Pullin, A., Frampton, G., Jong, R., Kohl, C., Livoreil, B., Lux, A., Pataki, G., Petrokofsky, G., Podhora, A., Saarikoski, H., Santamaria, L., Schindler, S., Sousa-Pinta, I., Vandewalle, M., Wittmer, H. 2016. Selecting appropriate methods of knowledge synthesis to inform biodiversity policy. Biodiversity and Conservation 25: 1285-1300.

Raiffa, H. 1968. Decision analysis: Introductory lecturers on choices under uncertainty. Reading, MA: Addison-Wesley.

Raymond, C.V., McCune, J.L., Rosner-Katz, H., Chadès, I., Schuster, R., Gilbert, B., Bennett, J. R. 2020. Combining species distribution models and value of information analysis for spatial allocation of conservation resources. Journal of Applied Ecology 57: 819-830.

Runge, M.C., Converse, S.J., Lyons, J.E. 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biological Conservation 144: 1214 -1223.

Rytwinski, T., Elmer, L.K., Taylor, J.J., Donaldson, L.A., Bennett, J.R., Smokorowski, K.E., Winegardner, A.K., Cooke, S.J. 2019. How effective are spawning-habitat creation or enhancement measures for substrate-spawning fish? A synthesis. Canadian Technical Report of Fisheries and Aquatic Sciences 3333: viii + 183 p.

Salafsky, N., Boshoven, J., Burivalova, Z., Dubois, N.S., Gomez, A., Johnson, A., Lee, A., Margoluis, R., Morrison, J., Muir, M., Pratt, S.C., Pullin, A.S., Salzer, D., Stewart, A., Sutherland, W.J., Wordley, C.F.R. 2019. Defining and using evidence in conservation practice. Conservation Science and Practice 2019; e27. https://doi.org/10.1111/csp2.27.

Schwartz, M.W., Cook, C.N., Pressey, R.L., Pullin, A.S., Runge, M.C., Salafsky, N., Sutherland, W.J., Williamson, M.A. 2018. Decision support frameworks and tools for conservation. Conservation Letters 11: 1–12. https://doi.org/10.1111/conl.12385.

Shojania, K.G., Sampson, M., Ansari, M.T., Ji, J., Doucette, S., Moher, D. 2007. How quickly do systematic reviews go out of date? A survival analysis. Annals of Internal Medicine 147: 224e33. https://doi.org/10.7326/0003-4819-147-4-200708210-00179.

Simmonds, M., Salanti, G., Ravaud, P., Higgins, H., McKenzie, J., Elliott, J., Living Systematic Review Network. 2017. Living systematic review: 3. Statistical methods for updating meta-analyses. Journal of Clinical Epidemiology 91: 38-46. doi: 10.1016/j.jclinepi.2017.08.008.

Smith, C.S., Howes, A.L., Price, B., McAlpine, C.A. 2007. Using a Bayesian belief network to predict suitable habitat of an endangered mammal–The Julia Creek dunnart (*Sminthopsis douglasi*). Biological Conservation 139 (3): 333–347.

Smith, M.L., Glass, G.V. 1977. Meta-analysis of psychotherapy outcome studies. American Psychologist 32(9): 752-760. doi: 10.1037//0003-066X.32.9.752.

Smith, S.L., Stern, N. 2011. Uncertainty in science and its role in climate policy. Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences 369(1956): 4818-4841.

Stewart, G., 2010. Meta-analysis in applied ecology. Biology Letters 6: 78–81.

Sutherland, W.J., Wordley, C.F. 2018. A fresh approach to evidence synthesis. Nature 558: 364–366.

Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M. 2004. The need for evidence-based conservation. Trends in Ecology & Evolution 19: 305-308.

Sutherland, W.J., Dicks, L.V., Ockendon, N., Petrovan, S.O., Smith, R.K., (editors). 2019. What Works in Conservation 2019. Open Book Publishers, Cambridge, UK.

Sutherland, W.J. Dias, M.P., Dicks, L.V., Doran, H., Entwistle, A.C., Fleishman, E., Gibbons, D.W., Hails, R., Hughes, A.C., Hughes, J., Kelman, R., Le Roux, X., LeAnstey, B., Lickorish, F.A., Maggs, L., Pearce-Higgins, J.W., Peck, L.S., Pettorelli, N., Pretty, J., Spalding, M.D., Tonneijck, F.H., Wentworth, J., Thorton, A. 2020. A horizon scan of emerging global biological conservation issues for 2020. Trends in Ecology & Evolution 35(1): 81-90.

Taylor, J.J., Rytwinski, T., Bennett, J.R., Smokorowski, K.E., Lapointe, N.W.R., Janusz, R., Clarke, K., Tonn, B., Walsh, J.C., Cooke, S.J. 2019. The effectiveness of spawning habitat creation or enhancement for substrate spawning temperate fish: a systematic review. Environmental Evidence 8:19. doi:10.1186/s13750-019-0162-6.

Tricco, A. C., Garritty, C. M., Boulos, L., Lockwood, C., Wilson, M., McGowan, J., McCaul, M., Hutton, B., Clement, F., Mittmann, N., Devane, D., Langlois, E.V., Abou-Setta, A.M., Houghton, C., Glenton, C., Kelly, S.E., Welch, V.A., LeBlanc, A., Wells, G.A., Pham, B., Lewin, S., Straus, S.E. 2020. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. Journal of clinical epidemiology, 126: 177-183.

Tulloch, A.I., Maloney, R.F., Joseph, L.N., Bennett, J.R., Di Fonzo, M.M., Probert, W.J., O'Connor, S.M., Densem, J.P., Possingham, H.P. 2015. Effect of risk aversion on prioritizing conservation projects. Conservation Biology 29(2): 513-524.

Tulloch, V.J., Possingham, H.P., Jupiter, S.D., Roelfsema, C., Tulloch, A.I. and Klein, C.J., 2013. Incorporating uncertainty associated with habitat data in marine reserve design. Biological Conservation 162: 41-51. https://doi.org/10.1016/j.biocon.2013.03.003.

Wright, A.D., Bernard, R.F., Mosher, B.A., O'Donnell, K.M., Braunagel, T., DiRenzo, G.V., Fleming, J., Shafer, C., Brand, A.B., Zipkin, E.F., Campbell Grant, E.H. 2020. Moving from decision to action in conservation science. Biological Conservation 249: 108698. https://doi.org/10.1016/j.biocon.2020.108698.

Box 1. A case study using the Canadian context.

In Canada, efforts to conduct and utilize systematic reviews for environmental management and conservation are still in their infancy. However, progress is ongoing. For example, institutions within the Canadian government, such Parks Canada, Environment and Climate Change Canada, and Fisheries and Oceans Canada (DFO), have recently begun integrating formal systematic reviews into their decision-making processes following guidelines developed by the Collaboration for Environmental Evidence (CEE) (2018). Highlighting one such case here, DFO’s Fish and Fish Habitat Protection Program (FFHPP) was seeking advice on best practices in habitat restoration and information on the effectiveness of restoration practices in regions of varying productivity and community compositions. To address this request, Taylor et al. (2019) conducted a systematic review (including a quantitative synthesis using formal meta-analytical methods) to assess the effectiveness of techniques currently used to create or enhance spawning habitat for substrate-spawning fish in temperate regions. This systematic review was conducted under the guidance of the CEE (2018), and as such allowed reviewers to identify the most relevant, and reliable (minimally biased) sources of information on the review topic. However, while the evidence base on the topic was relatively large, following such rigorous guidelines resulted in the exclusion of several studies from the systematic review because they were considered to be relatively low validity sources (i.e., susceptible to bias and/or had inadequate study designs). To gauge the amount of information gained from including available literature initially excluded from the Taylor et al. (2019) systematic review, a second (non-systematic) review (i.e., Rytwinski et al., 2019) was conducted to produce additional evidence for consideration in the agency’s formal science advisory process. These two documents formed the bases for a resulting Science Advisory Report (DFO, 2020) to provide science advice to DFO managers.

For further context, the systematic review (i.e., Taylor et al., 2019) used formal meta-analysis techniques to calculate effect sizes for various spawning habitat interventions. These effect sizes were based on the standardized mean difference between intervention and control groups (in this case, represented as a statistic known as Hedges' *g*), with individual studies weighted according to their standard error. To calculate such effect sizes, replication was required in study designs [i.e., >1 waterbody receiving a creation or enhancement of spawning habitat treatment and >1 waterbody not receiving the treatment, the control]. For the second review, to be inclusive as possible (i.e., allow inclusion of data sets that either lacked replication or that did not report variances or sample sizes for mean outcomes), Rytwinski et al. (2019) did not use formal meta-analytical methods. Instead, for any data set that had quantitative data [either a mean (number of replicates >1) or total count (n=1) for both the intervention and comparator group], they calculated the percent change in intervention effectiveness. Percent change is a more basic, less robust statistic not traditionally used in meta-analysis though it does provide some useful information that was otherwise excluded from the systematic review. In so doing, the number of data sets included in quantitative synthesis increased from 53 in the Taylor et al. (2019) systematic review to 228 in the Rytwinski et al. (2019) review. Within both the Rytwinski et al. (2019) review and the DFO (2020) report, comparisons between the two quantitative analyses were made, highlighting the similarities and differences in review conclusions (e.g., see Table 1 in DFO, 2020), but most importantly, both attempted to provide informative evidence with clear considerations for review limitations and caveats with respect to study validity. For instance, while the results from Rytwinski et al. (2019) supported the general findings from the systematic review, one of the most notable observations was that by adding the lower validity studies, there was evidence of increased uncertainty in the estimated effectiveness relative to the systematic review. Yet, this report did allow for the inclusion of a greater diversity of species and intervention types, leading to valuable products such as a curated database with a critical appraisal of included studies. As such, we highlight this case study as an example of how researchers can make use of the entire evidence base on a topic, attempting some form of analysis of the primary studies to make use of the entire evidence base on a topic, and making use of all review end products, so as long as this evidence is accompanied with appropriate considerations for study validity.

Box 2. Summary of recommendations.

<u>At the front-end of a systematic review</u>

Attempt a more rigorous scoping exercise that enables estimation of the size and reliability of the evidence base to identify the potential for ambiguity, bias, and the absence of additional information.

<u>To address *potential* ambiguity:</u>

- consider a systematic map as a starting point
- consider narrowing the scope of the review

<u>To address *potential* bias:</u>

- consider a different decision-making framework/tool

<u>To address *potential* absence of additional evidence:</u>

- consider broadening the scope of the review to capture more evidence
- consider broadening the review search to include openly accessible datasets to make use of additional data from non-target studies
- revisit other frameworks/tools
- proceed with the systematic review but communicate clearly with practitioners the limitations of the current evidence base

<u>At the back-end of an on-going or completed systematic review</u>

<u>To address ambiguity:</u>

- incorporate new evidence as it becomes available
- partition studies for pooling into comparable groupings (e.g., different outcomes) and conduct separate analyses
- investigate potential sources of heterogeneity
- attempt meta-analyses in multiple ways if two or more equally valid approaches are possible to see how results compare, presenting results within a range of uncertainty when results differ substantially

To address bias:

- investigate the influence of bias on the effect estimates when the evidence base has mixed reliability, and attempt to balance precision (making use of the entire evidence base) with minimizing biased systematic review results
- go beyond a simple narrative synthesis and attempt some form of analysis with a discussion of caveats and limitations, or make full use of systematic review end products (i.e., database of relevant evidence sources combined with detailed information on study validity) to help inform decisions when the evidence base consists largely of low reliability studies.

To address no additional evidence arising for a systematic review:

- re-visit problem formulation
- accept that fast action may be crucial, despite high uncertainty
- make use of additional tools that can help focus original field or lab research on the spatiotemporal scales that will inform management decisions (e.g., value of information theory, Bayesian belief networks)

Table 1. Terms and definitions used within this paper.

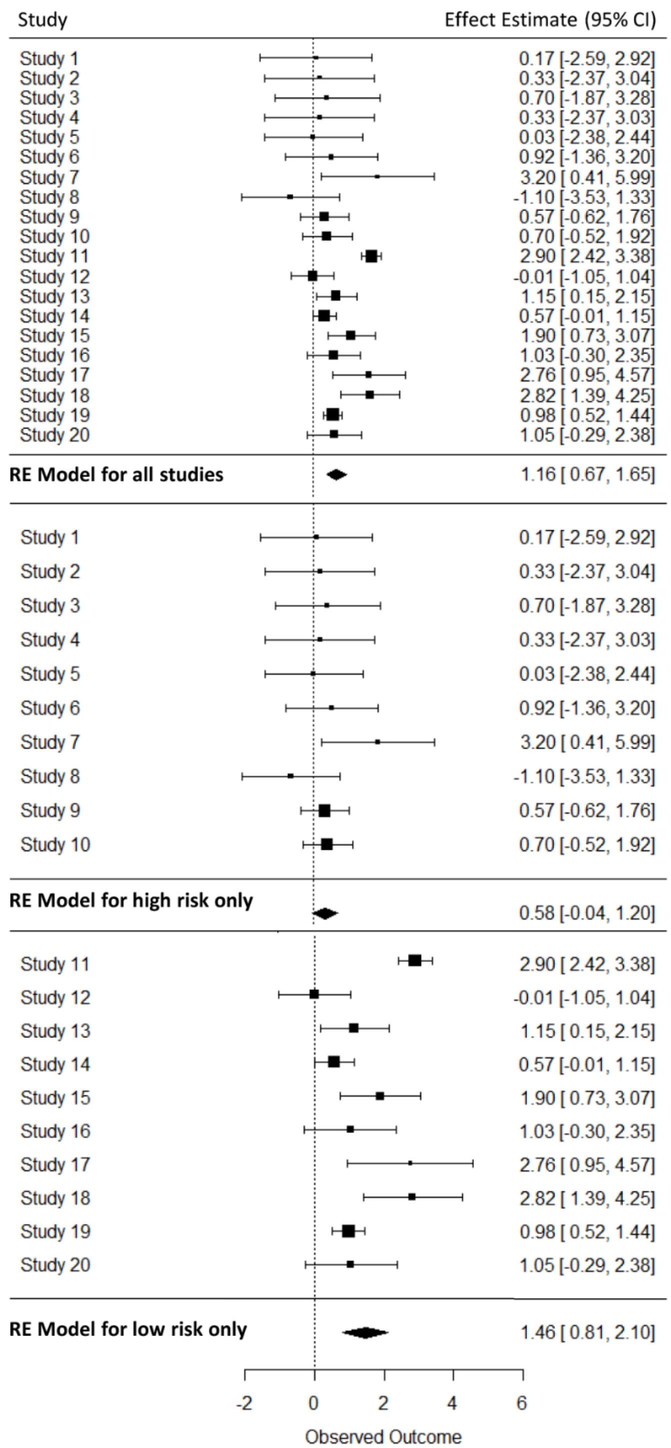
Term	Description
Systematic review process components	
Critical appraisal of study validity	An assessment of the comparative validity of the included studies requiring a number of decisions about the absolute and relative importance of different sources of bias and data validity elements common to environmental data (CEE, 2018). Ensures that all individual studies are objectively assessed for *internal validity* (reliability; is there potential for error and bias in the methodology employed to generate the study data) and *external validity* (generalisability; how transferable is the study to the context of the question). It can form a basis for the differential weighting of studies in later synthesis or partitioning of studies into subgroups for separate analyses (see *Critical appraisal of study validity (SRs)* in CEE, 2018).
Eligibility criteria	A predefined list of inclusion conditions (specified at the protocol stage) that determine which of the primary research studies identified in the searches are relevant for answering the review question; applied at the eligibility screening step of a systematic review (or systematic map) (CEE, 2018).
Decision-making frameworks/tools*	
Meta-analysis	A statistical tool used to combine the numerical results from across multiple studies to provide 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 (Stewart, 2010).
Multiple expert consultation with Delphi method	With the help of a coordination team or a facilitator, this method combines the knowledge of multiple, carefully selected experts into either quantitative or qualitative assessments, using a formal consensus on the question (described and reviewed by Mukherjee et al., 2016; Pullin et al., 2016; Dicks et al., 2017).

Structured decision making	A well-defined method for analyzing a decision by breaking it into components including the objectives, possible actions, and models linking actions to objectives. It relies on the integration of scientific information and stakeholder values to develop solution strategies, and as such, provides inclusion and transparency throughout the decision-making process (Bower et al., 2018). It is organized into clearly delineated steps that formulate the decision-making framework (see Gregory et al., 2012 for details on each step, and Dicks et al., 2017 and Schwartz et al., 2018 for details on framework functionality and comparisons).
Systematic conservation prioritization	Refers to a broad set of tools for quantitatively ranking conservation actions to maximize outcomes given limited resources; all of these tools share a similar structure (Margules and Pressey, 2000). It is most suited to problems where options are chosen based on trade-offs among attributes that are quantified using consistent measurements across all units (see Bower et al., 2018; Schwartz et al., 2018).
Systematic map	A form of evidence synthesis that aims to provide an accurate description of the evidence base relating to a particular question where methods are specified a priori in a protocol. Although procedurally similar to a systematic review, systematic maps do not aim to provide a quantitative or qualitative answer to a particular question, but instead, an overview of research that has been undertaken (Haddaway et al., 2016; James et al., 2016). Reviewers use predefined methods to minimize bias in the way the evidence is identified and selected. A descriptive overview of the evidence base is developed that could inform further research and synthesis (e.g., by revealing knowledge gaps and identifying more specific questions suitable for Systematic Review) (CEE, 2018).
Systematic review	A highly structured form of evidence synthesis where methods are specified a priori in a protocol. The goal of a systematic review is to answer a specific question as precisely as possible in an unbiased way. The process includes collating all relevant evidence and critical appraisal of the included evidence. Reviewers use predefined methods to identify risks of bias in the evidence itself, and to minimise bias in the way evidence is identified and selected, and thus provide reliable findings that could inform decision making. May include a quantitative synthesis of the included evidence to improve precision (Pullin et al., 2016; CEE, 2018).

Issues arising from a systematic review	
Evidentiary ambiguity	A form of uncertainty, whereby the uncertainty of the impact(s) cannot be quantified via probabilities; acknowledging that this sometimes also reflects quantifiable imprecision (another form of uncertainty). Therefore, for the purpose of this paper, we describe it as ambiguity arising from the evidence base and results of a systematic review, whether due to quantified or unquantifiable uncertainty, potentially making the appropriate decision or management response unclear because it can be understood in more than one way.
Evidentiary bias	Bias as a result of a systematic error; a systematic deviation in study results from their true value (CEE, 2018). When a systematic review is based on biased evidence, the results of the quantitative synthesis of a systematic review will also be incorrect, leading to misleading conclusions.
Absence of additional evidence	The absence of any *new* evidence arising from a systematic review (opposed to the absence of any evidence at all); acknowledging that no additional *evidence* may still amount to important additional *information*.

*This is not a full list of frameworks/tools for evidence-based decision-making in applied environmental science; listed here are terms/phrases that are referred to in this paper. For more comprehensive lists and descriptions/comparisons of decision-making frameworks/tools, see Dicks et al., 2014; Bilotta et al., 2015; Haddaway et al., 2015; Pullin et al., 2016; Cook et al., 2017; Dicks et al., 2017; Bower et al., 2018; Schwartz et al., 2018.

Fig. 1. An example of a stratified forest plot (using hypothetical data) displaying overall effect size estimates of the intervention effect from meta-analyses (using a random effect model=RE) based on: all studies regardless of risk of bias (top panel); only studies at high risk of bias i.e., low validity studies (middle panel); and only studies at low risk of bias i.e., high validity studies (bottom panel). In the example provided, although the relative magnitude of intervention effectiveness appears to be influenced by study validity, with higher estimated mean increases in outcomes for the analysis based on studies with lower susceptibility to bias (bottom panel), all analyses trend towards the same conclusion (i.e., that there is an estimated positive effect of the intervention regardless of whether lower validity studies are included).


Study
Effect Estimate (95% CI)
Study 1 0.17 [-2.59, 2.92]
Study 2 0.33 [-2.37, 3.04]
Study 3 0.70 [-1.87, 3.28]
Study 4 0.33 [-2.37, 3.03]
Study 5 0.03 [-2.38, 2.44]
Study 6 0.92 [-1.36, 3.20]
Study 7 3.20 [0.41, 5.99]
Study 8 -1.10 [-3.53, 1.33]
Study 9 0.57 [-0.62, 1.76]
Study 10 0.70 [-0.52, 1.92]
Study 11 2.90 [2.42, 3.38]
Study 12 -0.01 [-1.05, 1.04]
Study 13 1.15 [0.15, 2.15]
Study 14 0.57 [-0.01, 1.15]
Study 15 1.90 [0.73, 3.07]
Study 16 1.03 [-0.30, 2.35]
Study 17 2.76 [0.95, 4.57]
Study 18 2.82 [1.39, 4.25]
Study 19 0.98 [0.52, 1.44]
Study 20 1.05 [-0.29, 2.38]
RE Model for all studies 1.16 [0.67, 1.65]
Study 1 0.17 [-2.59, 2.92]
Study 2 0.33 [-2.37, 3.04]
Study 3 0.70 [-1.87, 3.28]
Study 4 0.33 [-2.37, 3.03]
Study 5 0.03 [-2.38, 2.44]
Study 6 0.92 [-1.36, 3.20]
Study 7 3.20 [0.41, 5.99]
Study 8 -1.10 [-3.53, 1.33]
Study 9 0.57 [-0.62, 1.76]
Study 10 0.70 [-0.52, 1.92]
RE Model for high risk only 0.58 [-0.04, 1.20]
Study 11 2.90 [2.42, 3.38]
Study 12 -0.01 [-1.05, 1.04]
Study 13 1.15 [0.15, 2.15]
Study 14 0.57 [-0.01, 1.15]
Study 15 1.90 [0.73, 3.07]
Study 16 1.03 [-0.30, 2.35]
Study 17 2.76 [0.95, 4.57]
Study 18 2.82 [1.39, 4.25]
Study 19 0.98 [0.52, 1.44]
Study 20 1.05 [-0.29, 2.38]
RE Model for low risk only 1.46 [0.81, 2.10]
-2
0
2
4
6
Observed Outcome