

Analyzing cause and effect in environmental assessments: using weighted evidence from the literature

Authors: Norris, R. H, Webb, J. A, Nichols, S. J, Stewardson, M. J, and Harrison, E. T

Source: Freshwater Science, 31(1) : 5-21

Published By: Society for Freshwater Science

URL: <https://doi.org/10.1899/11-027.1>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

PERSPECTIVES

This section of the journal is for the expression of new ideas, points of view, and comments on topics of interest to aquatic scientists. The editorial board invites new and original papers as well as comments on items already published in Freshwater Science (formerly J-NABS). Format and style may be less formal than conventional research papers; massive data sets are not appropriate. Speculation is welcome if it is likely to stimulate worthwhile discussion. Alternative points of view should be instructive rather than merely contradictory or argumentative. All submissions will receive the usual reviews and editorial assessments.

Analyzing cause and effect in environmental assessments: using weighted evidence from the literature

R. H. Norris^{1,4}, J. A. Webb^{2,5}, S. J. Nichols^{1,6}, M. J. Stewardson^{3,7}, AND
E. T. Harrison^{1,8}

¹ eWater Cooperative Research Centre, Institute for Applied Ecology, University of Canberra, Australian Capital Territory 2601, Australia

² eWater Cooperative Research Centre, Department of Resource Management and Geography, The University of Melbourne, Victoria 3010, Australia

³ eWater Cooperative Research Centre, Department of Infrastructure Engineering, The University of Melbourne, Victoria 3010, Australia

Abstract. Sound decision making in environmental research and management requires an understanding of causal relationships between stressors and ecological responses. However, demonstrating cause–effect relationships in natural systems is challenging because of difficulties with natural variability, performing experiments, lack of replication, and the presence of confounding influences. Thus, even the best-designed study may not establish causality. We describe a method that uses evidence available in the extensive published ecological literature to assess support for cause–effect hypotheses in environmental investigations. Our method, called *Eco Evidence*, is a form of causal criteria analysis—a technique developed by epidemiologists in the 1960s—who faced similar difficulties in attributing causation. The *Eco Evidence* method is an 8-step process in which the user conducts a systematic review of the evidence for one or more cause–effect hypotheses to assess the level of support for an overall question. In contrast to causal criteria analyses in epidemiology, users of *Eco Evidence* use a subset of criteria most relevant to environmental investigations and weight each piece of evidence according to its study design. Stronger studies contribute more to the assessment of causality, but weaker evidence is not discarded. This feature is important because environmental evidence is often scarce. The outputs of the analysis are a guide to the strength of evidence for or against the cause–effect hypotheses. They strengthen confidence in the conclusions drawn from that evidence, but cannot ever prove causality. They also indicate situations where knowledge gaps signify insufficient evidence to reach a conclusion. The method is supported by the freely available *Eco Evidence* software package, which produces a standard report, maximizing the transparency and repeatability of any assessment. Environmental science has lagged behind other disciplines in systematic assessment of evidence to improve research and management. Using the *Eco Evidence* method, environmental scientists can better use the extensive published literature to guide evidence-based decisions and undertake transparent assessments of ecological cause and effect.

Key words: causal criteria analysis, causality, causal inference, weak inference, *Eco Evidence*, systematic review, environmental assessment.

⁴ Deceased

⁵ E-mail address: angus.webb@unimelb.edu.au

⁶ To whom correspondence should be addressed. E-mail: sue.nichols@canberra.edu.au

⁷ E-mail addresses: mjstew@unimelb.edu.au

⁸ evan.harrison@canberra.edu.au

Do dams on rivers cause changes in fish assemblages? More specifically, will building a new dam negatively affect the endangered native species found at the proposed dam site? These types of general and specific questions concerning human impacts on

the environment are being asked of environmental scientists every day. How much confidence can we have in our answers? It may be possible to see statistical associations between apparent stressors and indicators of environmental degradation, but reaching a conclusion with an acceptable level of confidence that one thing actually *causes* another is challenging in environmental research. However, causal understanding in environmental science is vital for answering general questions of causality in natural environments and for addressing specific instances of environmental degradation.

Weak inference in environmental sciences

Environmental investigations often are carried out in situations where robust conclusions concerning hypothesized causes and effects are difficult to draw (Beyers 1998). For example, we might observe a difference between the fish assemblages upstream and downstream of a particular dam, but several potential explanations may exist for this observation other than the dam itself. Difficulty in inferring the most likely cause of such differences stems from several sources. First, studies are often observational, and from a study-design point of view, any treatments cannot be randomly allocated (Beyers 1998, Johnson 2002). To extend the example, dams are not placed randomly within stream networks but are situated at locations where river and valley hydro-morphology deliver the best outcomes in terms of efficient water storage and supply. Thus, differences in the fish assemblages could be caused by the differences in river morphology that affected the choice of dam location and may have existed before the dam was built. The study is happening after the event, control sites may not be available, and replication cannot be imposed.

This example demonstrates the main factors that can lead to a study design that lacks one or more characteristics that would allow us to infer a cause-effect relationship with reasonable confidence (Downes et al. 2002). Often, no data exist that describe the putatively disturbed location before development, allocating control locations is difficult, confounding environmental factors exist, and replication is insufficient in these naturally variable environments. In many situations, a lack of *before* data is inevitable because the purpose of the investigation is to determine ecological effects of prior or concurrent developments. Control or reference locations may not be available, particularly for large-scale disturbances or systems. Environmental gradients between disturbed and control sites and factors, such as rainfall, temperature, and latitudinal

effects, can confound the interpretation of any observed ecological difference. Inference also is weakened by insufficient replication in space and time (Johnson 2002). At the level of the treatment, multiple independent, potentially disturbed locations are rarely examined within the same investigation. Thus, inherent differences between the potentially disturbed and control locations are not accounted for, and the chances of confounding are greater, particularly because natural environments often exhibit great variation among locations. Replication of control locations defines an envelope of undeveloped conditions against which to compare the impact location, thus reducing the likelihood of spurious conclusions. Again, however, replicate control locations often are not available for environmental investigations.

Provision of before-development data, multiple control locations, and appropriate replication are stipulated as the minimum data requirements for inferring human impacts on the environment (Green 1979, Downes et al. 2002). A single study with these characteristics might allow us to reach a confident conclusion if it were sufficiently powerful to detect changes considered ecologically significant. However, the difficulties described above mean that, in many studies, these requirements are not fully met (Norris et al. 2005). Thus, one could argue that very few studies of human impacts on the environment can provide a severe (*sensu* Popper 1983) test of an hypothesized cause-effect relation, and that strong inference (*sensu* Platt 1964) is seldom possible with individual studies. Thus, novel (and robust) methods are required to assess cause-effect hypotheses about human impacts, particularly if a legal challenge is to be made or if management must balance ecological health with economic or social considerations.

Causal criteria and multiple lines of evidence

If one can seldom infer cause and effect from individual ecological studies, then additional evidence is needed (Downes et al. 2002). The evidence might be from sources as wide-ranging as repeated studies of the same hypothesized cause-effect relation in different environments and with different study designs and methods, experimental results from small-scale manipulations in the laboratory or field, or evidence of the hypothesized causal agent within the target organism (e.g., body burden of heavy metals in fish near mine sites). Individually, none of these types of evidence may be convincing, but together they may provide numerous lines of evidence (*sensu* Norris et al. 2005) that amount to a

TABLE 1. The causal criteria defined by Hill (1965) for use in epidemiology. These criteria were developed from criteria originally defined in the US Surgeon General's report on the health effects of smoking (USDHEW 1964). Subsequent users have tended to concentrate on smaller subsets of the 9 criteria.

Criterion	Definition
Strength of association ^{abc}	Increase in disease incidence associated with exposure to the risk factor. Measured as the likelihood (i.e., increased risk) of developing the disease when exposed to the risk factor.
Consistency of association ^{abc}	An association between the hypothesized cause and effect has been observed repeatedly in a variety of places, circumstances, and times.
Specificity of association ^{ac}	Refers to the case where either: 1) the effect is found only in association with the hypothesized cause or 2) only one effect is associated with the cause.
Temporality ^{ac}	The hypothesized cause is observed to precede effect.
Biological gradient (dose-response) ^{bcd}	A sensible relationship exists between the dose of the hypothesized cause and the level of response observed.
Biological plausibility ^{bc}	A conceptually sensible biological explanation exists for the relationship between cause and effect.
Coherence ^a	Support for the cause-effect relation is provided by known facts including laboratory results or patterns in association between cause and effect between populations.
Experiment ^c	Results from experimental studies support the cause-effect relationship. This criterion applies only to situations where the hypothesized cause is manipulated experimentally.
Analogy	The hypothesized cause-effect relation can be argued on the basis of similarity with an established cause and effect.

^a Original criteria defined by USDHEW (1964)

^b Criteria quantitatively identified by Weed and Gorelic (1996) as being the most commonly used in epidemiology

^c Principal causal criteria identified by Adams (2005)

^d Weed and Gorelic (1996) and Weed (1997) used the name 'dose-response' for this criterion

powerful argument for causality. Intentionally or otherwise, environmental researchers often seek to strengthen their arguments for causality by informally including other lines of evidence from the literature in the discussion sections of their research papers (reviewed in Downes et al. 2002). However, until now, we have lacked a rigorous framework for synthesizing these lines of evidence.

Here, we introduce a method that considers the evidence from many studies to transparently assess the level of support for questions of cause and effect. The method, called *Eco Evidence*, is a form of causal criteria analysis. Causal criteria were developed by epidemiologists in the 1960s for assessing cause-effect hypotheses in the face of weak evidence (Weed 1997, Tugwell and Haynes 2006). The causal criteria are a checklist. Each hypothesized cause-effect relation is assessed against a series of criteria that have as their philosophical basis the Henle and Koch postulates for inferring causes of disease (Evans 1976). The best-known set of epidemiological causal criteria was developed by Hill (1965), but exactly which criteria are adopted for particular studies is variable. The most commonly used causal criteria in epidemiology are detailed in Table 1 (Weed and Gorelic 1996).

Fox (1991) was the first to suggest that the epidemiological approach for assessing causation would be suitable for use in the environmental

sciences, but little debate or testing of appropriate criteria for use in this area has ensued (Beyers 1998, Downes et al. 2002, Suter et al. 2002, Adams 2005, Plowright et al. 2008). Thus, consistency and clarity in the criteria used are lacking among the relatively few existing case studies (Lowell et al. 2000, Cormier et al. 2002, Downes et al. 2002, Collier 2003 and case studies therein, Fabricius and De'Ath 2004, Burkhardt-Holm and Scheurer 2007, Haake et al. 2010, Wiseman et al. 2010). Of these studies, all but Downes et al. (2002) analyzed existing data or new studies at a particular site. Moreover, although Fox (1991) and Suter et al. (2010) provided methods to quantify fulfillment of individual causal criteria, only Norris et al. (2005) provided a quantitative method to combine information across criteria to assess the overall level of support for causality. The method used by Norris et al. (2005) also is generally applicable to the range of studies undertaken in environmental science, including investigations of human impacts and more general investigations of cause and effect in natural environments.

Eco Evidence is a more fully developed version of the method published by Norris et al. (2005) and is supported by freely available software. It incorporates concepts suggested by Suter et al. (2002, 2010), Adams (2005), and Downes et al. (2002). The method operates within the conjecture-refutation model of scientific progress familiar to most researchers and can identify

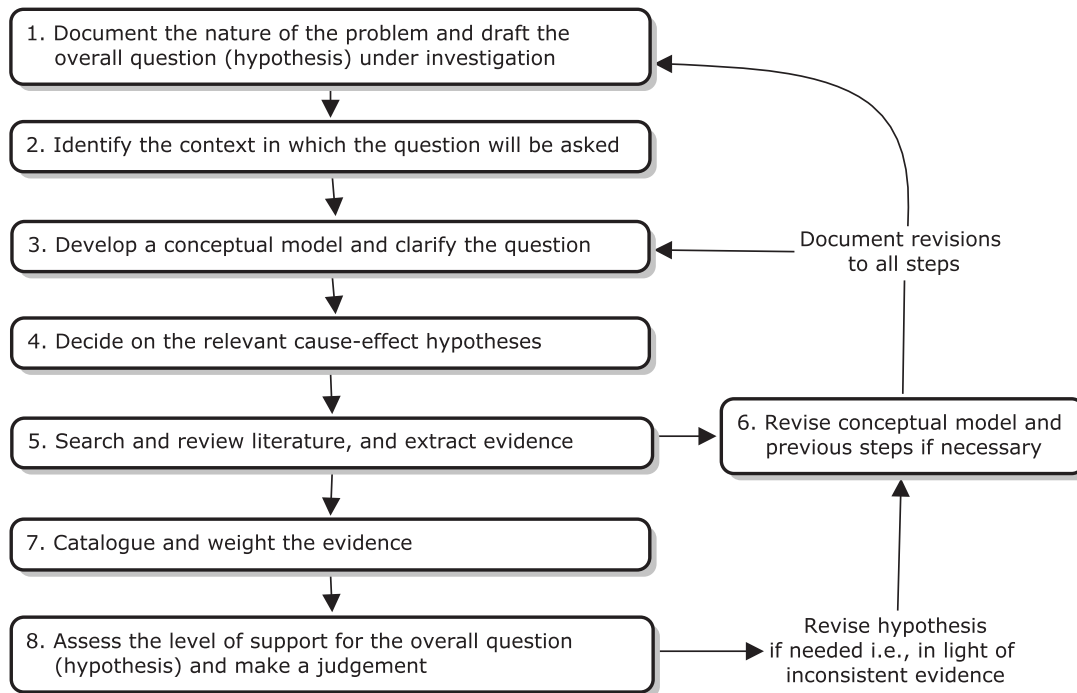


FIG. 1. Steps in the Eco Evidence framework. Adapted from Nichols et al. (2011).

knowledge gaps. It centers on systematic review of the extensive pool of existing scientific literature to assess transparently the level of support for cause-effect hypotheses. Each published study relevant to the topic in question is weighted by its ability to contribute to the argument for causality. Stronger studies contribute more to the assessment, but weaker evidence is not discarded. This feature is important because environmental evidence is often scarce. All studies are considered collectively against one or more hypotheses, and the specific cause-effect hypotheses assessed depend on the specific ecological question at hand (also see Suter et al. 2002). Each hypothesis is then evaluated against several causal criteria.

Method: the Eco Evidence Analysis Framework

Eco Evidence is an 8-step framework and method (Fig. 1) that is presented in full detail by Nichols et al. (2011). The method is a form of systematic review of the scientific literature. Thus, in our paper, the user is called the *reviewer*. Steps 1 to 4 and 6 of the method can be loosely described as *problem formulation*. The reviewer documents the nature of the problem under investigation and formulates an overall question (hypothesis), identifies the context in which the question will be asked, develops a conceptual model of the problem, and documents the hypothesized

cause-effect relationships within the overall question that will be tested. Step 5 consists of the literature search and systematic review. The reviewer extracts and collates evidence from relevant literature. The reviewer then reconsiders the conceptual model in the light of the literature review, and decides whether Steps 1–4 should be revised (Step 6). In Steps 7 and 8, the reviewer weights, combines, and considers the evidence to assess the level of support for and against the individual cause-effect hypotheses identified at Step 4. These results are then assessed collectively to inform an overall finding in relation to the overall question developed at Step 1 (Nichols et al. 2011). The 8 steps are described in detail below and are illustrated through a simplified case study that examines the effect of fine sediment on stream invertebrates (Harrison 2010). For case studies presented in full detail, readers are directed to Harrison (2010) and Greet et al. (2011).

The Eco Evidence framework is supported by a software package (eWater CRC 2010) that consists of an online database for storing evidence extracted from publications and a desktop analysis tool with a user interface to guide the reviewer through the 8-step causal inference process (Wealands et al. 2009, Webb et al., in press b). The software also produces a full report of the analysis, which details all inputs for the 8 steps, the evidence used in the assessment, and how it was weighted and interpreted. This report provides full transparency and repeatability for any Eco

Evidence analysis. Version 1.0 Eco Evidence was released in September 2011 and is available free of charge from www.toolkit.net.au/tools/eco-evidence.

Step 1: Document the nature of the problem and draft the overall question (hypothesis) under investigation

The 1st step is to document the nature of the problem under investigation and draft the overall question (hypothesis), hereafter referred to as *The Question*. This question may be specific to a particular problem (e.g., Will a proposed new dam reduce native fish abundance in one particular river?) or more general (e.g., How are native fish affected by river damming?). The reviewer records the hypothesized causal agent(s), the potential effect(s), and considers their timing and magnitude. In the case study (Harrison 2010), land-clearing practices induce soil erosion and increase the delivery of fine sediment to streams, which is considered a major detriment to the ecological condition of rivers. The Question was: How does sediment in streams affect macroinvertebrate assemblages? Harrison (2010) refined this question into 3 subquestions that were each tested using the Eco Evidence method:

1. Does macroinvertebrate community structure change as a result of accumulation of fine sediment within the stream bed?
2. Does a threshold level (i.e., % stream bed covered) of accumulation of fine sediment exist above which macroinvertebrate community structure changes?
3. Does macroinvertebrate community structure change in response to increased transport of fine sediment through a river reach?

Step 2: Identify the context in which The Question will be asked

At Step 2, the reviewer describes the context. Will the review be limited to an environment of a particular type or be more general? Is a geographic restriction on studies appropriate? Is the review concerned only with causal agents that originate from one source, or are all sources of a causal agent relevant? The context is used later to help identify published studies relevant to the assessment. The effects of human-induced increases in fine sediments in streams were the primary interest for the Harrison (2010) case study. Stream type was not restricted based on geography or climate, but the review was constrained to studies that considered effects on macroinvertebrates of increased transport or accumulation of fine sediment in streams caused by

human activities or experimental manipulations of sediment.

Step 3: Develop a conceptual model and clarify The Question

Step 3 is very important, because a well considered conceptual model provides the basis for the remainder of the causal assessment (Nichols et al. 2011). It provides transparency by clearly displaying the hypothesized causal relationships in the context of the problem under investigation. Where possible, such linkages should be process-based rather than simple empirical associations (Cormier et al. 2010). The conceptual model should include potential confounding variables, and other potential causal agents. Ideally, all such variables should be assessed within the analysis to reduce the likelihood of reaching a spurious conclusion. However, logistic constraints may mean that the reviewer has to prioritize variables for assessment to those most likely to affect the causal interpretation. Although not compulsory, graphical conceptual models are useful for illustrating the mechanisms that explain the hypothesized causal relationships. The reviewer may develop his or her own conceptual model, or use (or modify) an existing model deemed suitable. The library of conceptual models being developed through the US Environmental Protection Agency's (EPA) Interactive Causal Diagram tool (www.epa.gov/caddis/cd_icds_intro) would be an appropriate source of existing models. By specifying the hypothesized cause-effect linkages, the conceptual model will heavily guide the choice of relevant studies at Step 5 (Greet et al. 2011). Harrison (2010) developed a conceptual model of the effects of increased fine sediment transport and accumulation on in-stream habitat and macroinvertebrate community structure (Fig. 2) and then used it to identify specific cause-effect hypotheses for assessment.

Step 4: Decide on the relevant cause-effect hypotheses

At Step 4, based upon the conceptual model, the reviewer refines the preliminary list of hypothesized causal agents and effects. The conceptual model identifies multiple hypothesized cause-effect links within The Question that need to be prioritized for assessment. Causal agents and potential effects must be quantifiable (e.g., water temperature, population abundance). In the case study (Harrison 2010), the 3 subquestions grouped the effects within 3 categories: general indicators of community structure, number and abundance of fine-sediment-sensitive taxa, and number and abundance of fine-sediment-tolerant taxa

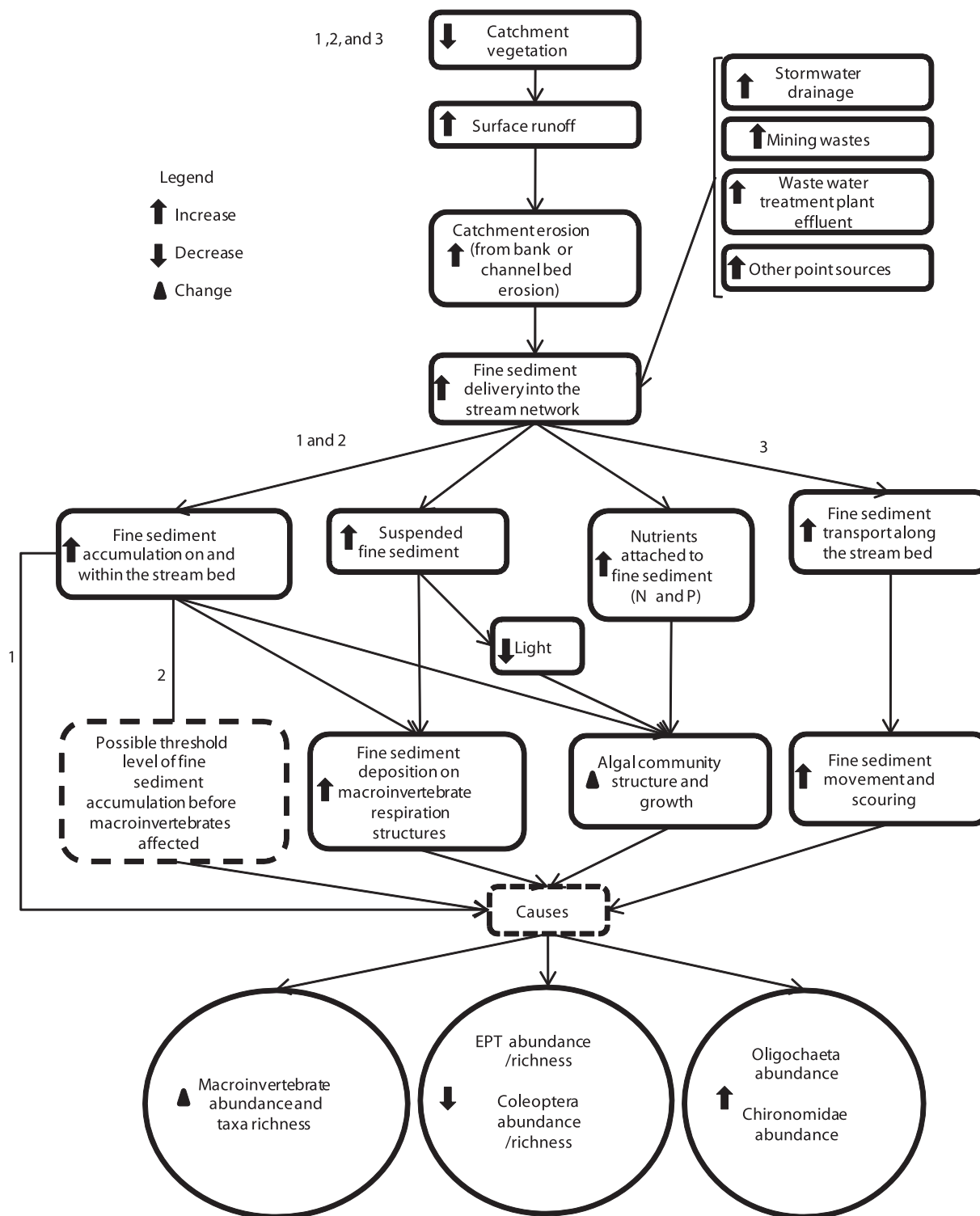


FIG. 2. Conceptual model of the effects of increased fine-sediment transport and accumulation on in-stream habitat and macroinvertebrate community structure. When fine sediment from gully and bank erosion enters a stream and accumulates within the stream bed, especially in reaches with low gradients, it reduces interstitial spaces that macroinvertebrates use for refuge, reduces periphyton growth, and increases deposition on macroinvertebrate respiration structures. These effects are expected to result in a change in macroinvertebrate abundance and taxon richness (Table 2). Numbers on the figure indicate the relevance of particular causal pathways to the 3 subquestions identified in the text. EPT = Ephemeroptera, Plecoptera, Trichoptera taxa.

TABLE 2. Expected effects on macroinvertebrates of an increase in accumulation or transport of fine sediment within the stream bed. EPT = Ephemeroptera, Plecoptera, Trichoptera taxa.

Quantifiable effects	Ways in which the effects are quantified
General indicators of community structure	Change in macroinvertebrate taxon richness Change in macroinvertebrate abundance
Changes in taxa sensitive to fine-sediment accumulation	Decreased EPT richness Decreased EPT abundance Decreased Coleoptera richness Decreased Coleoptera abundance
Changes in taxa tolerant of fine sediment	Increased Oligochaeta abundance Increased Chironomidae abundance

(Table 2). Consistent with the original questions asked (Step 1), the quantifiable causal agents were an increase in fine-sediment accumulation and an increase in fine-sediment transport through reaches.

Step 5: Search and review literature, and extract evidence

Step 5 involves the literature review. For repeatability and transparency, the Eco Evidence framework requires the reviewer to document the method used to search the literature (e.g., databases searched, search terms used), and then justify the inclusion (or exclusion) of all the studies initially delivered from the search. A study's title and abstract generally provide the information the reviewer requires to establish if the study is relevant. Justification for relevance could include, e.g., a combination of geographical proximity, similar environmental characteristics, and similar causal agents. However, relevant studies need not be drawn only from systems completely similar to the focus of the assessment. Moreover, laboratory and other small-scale manipulative studies can be particularly relevant because they are less likely to suffer from confounding. When selecting relevant studies, reviewers also may use matching and restriction-type approaches (Rothman et al. 2008) to reduce the likelihood of confounding. Harrison (2010) used the Cambridge Scientific Abstracts and Thompson Institute of Scientific Investigation (ISI) Web of Knowledge, and he obtained further studies cited in the papers found in the initial searches. The search used the keyword phrase ("sediment" OR "sedimentation") AND "macroinvertebrates". As described in Step 2, field studies were restricted to those that dealt with human-induced sediment accumulation. The search yielded 48 studies that were considered relevant to question 1, 5 studies relevant to question 2, and 3 studies relevant to question 3.

Once a study has been designated as relevant to the analysis, evidence in the study must be extracted. The *evidence item* extracted from the relevant study has 2

parts. First, the reviewer must determine whether an association exists that is consistent with the cause-effect hypothesis being tested and the nature of that association, including whether it presents as a dose-response relationship. A reviewer often may use statistical significance to assess whether an association exists. However, this rule is not general because it: 1) precludes the possibility of using studies in which statistical significance was not assessed, and 2) may lead to an inappropriate assessment of an ecologically irrelevant association in a study with very high replication. Second, the reviewer must record the type of study design used, choosing from the list of broad design categories in Table 3, and the number of independent sampling units. In a factorial design, sampling units are designated as either *Control* or *Impact*. For gradient-based design, total replication is recorded. Detailed guidance on assigning the type of study and on counting of sampling units was provided by Nichols et al. (2011). Regardless of the specific objectives of the study being assessed, the power of the study design (i.e., type of study and number of independent sampling units) will be determined by the relationship of the study design to the causal agent of concern to the reviewer. These 2 objectives often will coincide, and it will be appropriate to record the design and replication as used by the study authors in statistical analyses, but occasions will arise when the Eco Evidence analysis is being conducted on a causal agent that was not the focus of the original study, or which is being assessed at a different scale. In such cases, the design or replication recorded for the Eco Evidence analysis may differ from that used by the study authors to assess their specific objectives. We are not suggesting that the reviewer should reanalyze the findings of the study being assessed, nor are we passing any judgment on the appropriateness of the analyses undertaken for that study. The reviewer should record the reasons for such choices of design or replication as part of the justification for the relevance

TABLE 3. Weights applied to study types and the number of control/reference and impact/treatment sampling units (Nichols et al. 2011). B = before, A = after, C = control, R = reference, I = impact, M = multiple. See text for explanation of "Weight". Overall evidence weight is the sum of design weight and replication weight (factorial design or gradient response).

Study design component	Weight
Study design type	
After impact only	1
Reference/control vs impact with no before data	2
Before vs after with no reference/control location(s)	2
Gradient response model	3
BACI, BARI, MBACI, or beyond MBACI	4
Replication of factorial designs	
Number of reference/control sampling units	
0	0
1	2
>1	3
Number of impact/treatment sampling units	
1	0
2	2
>2	3
Replication of gradient-response models	
<4	0
4	2
5	4
>5	6

of the evidence item to their analysis. The information on design and replication is used to weight the evidence at Step 7.

Step 6: Revise conceptual model and previous steps if necessary

At Step 6, the reviewer decides if the conceptual model must be revised. During the literature review, the reviewer may learn about new potential causes, effects, or linkages relevant to the question being asked or may decide that previously listed causes and effects are irrelevant. Alternatively, the studies found may mean that individual hypotheses must be approached indirectly (i.e., to infer that A causes C, we must show separately that A causes B and B causes C; e.g., Greet et al. 2011). Harrison (2010) found no need to revise the conceptual model in light of the literature review. For an example of a conceptual model revised at Step 6, see Greet et al. (2011). A revision at this step entails an iterative return to previous steps (Fig. 1).

Step 7: Catalogue and weight the evidence

In Step 7, the reviewer records the amount and strength of evidence for and against each cause–effect

linkage under investigation. In the Eco Evidence framework, the reviewer uses a rule-based approach to weight individual studies. This approach to weighting evidence markedly distinguishes Eco Evidence from other existing applications of causal inference in either epidemiology or environmental science. Information on study design and replication of sampling units is weighted, and these 2 weights are summed to provide an overall weight for each individual study (Table 3).

The philosophy adopted in Eco Evidence is that studies that better account for environmental variability or error (e.g., before–after control–impact [BACI] designs; Green 1979, Underwood 1997) should carry more weight in the overall analysis than those with less robust designs (Norris et al. 2005, Nichols et al. 2011). Inclusion of control or reference sampling unit(s) improves a study's inferential power, as do provision of data from before the hypothesized disturbance or use of gradient-response models designed to quantify relationships between hypothesized cause and effect (see Downes et al. 2002). Studies in which several replicates have been included can provide an estimate of variability around normal conditions, a feature that adds weight to the findings because any difference detected between treatment and control was more likely to have been caused by the treatment (Downes et al. 2002).

The reviewer sums the component weights derived from study design type and replication for each study to determine an overall study weight, which can range between 1 and 10. These default weights (Table 3) were derived from numerous trials and extensive consultation with ecologists (discussed later in our paper). The reviewer can modify the default weights, but any such changes should be documented and justified. For example, a study of floodplain geomorphological processes (Grove et al., in press) redefined study weights reasoning that studies with 'before' data are simply not possible with the temporal scales of floodplain formation (i.e., 1000s of years). In the case study, Harrison (2010) used the default study weights (Table 3) to weight the evidence.

Step 8: Assess the level of support for the overall question (hypothesis) and make a judgment

In Step 8, the reviewer uses the summed study weights when evaluating support for the cause–effect hypotheses and collectively considers these results to reach a conclusion concerning The Question. Eco Evidence uses 6 causal criteria that we think most practically apply to environmental questions, but not

TABLE 4. Causal criteria adopted in the 8-step Eco Evidence framework for application in environmental sciences.

Causal criterion	Description	Basis for selection
Plausibility	A plausible mechanism (e.g., biochemical reaction) that could explain the relationship between the causal agent and the potential effect	This criterion is addressed by the conceptual model. A plausible conceptual model is a necessary step for any further analysis because it sets the bounds of the literature review, clearly displaying the causal relationships considered.
Evidence of response	An association between the causal agent and potential effect	This criterion is similar to the US Environmental Protection Agency's (EPA) criterion of co-occurrence, which assesses whether the hypothesized cause and effect are found at the same site (Suter et al. 2002). Studies are most easily classified as contributing to this criterion when there is a statistically significant difference among treatment levels.
Dose-response ^a	The association between causal agent and potential effect is in the form of a dose-response curve; the relationship normally would be monotonic	This criterion is analogous to Hill's (1965) biological gradient criterion, but we have avoided use of the term gradient to avoid confusion with gradient-based study designs, which is where most evidence of this type would be observed. Weed and Gorelic (1996) and Weed (1997) named this criterion dose-response, and we have adopted this name. However, we accept that for applications in Eco Evidence, true doses (i.e., a stressor being taken up by the organism) would occur only rarely.
Consistency of association	The potential effect occurs in the presence of the causal agent in all or almost all of the studies	Because of the frequent confounding and low statistical power that besets environmental studies, the presence of small numbers of nonsignificant or even contrary results does not necessarily indicate a lack of consistency (which is why we have incorporated criterion thresholds). In this respect, the criterion is less stringent than specificity of association (Hill 1965), which requires the effect to be seen always in the presence of the cause.
Evidence of stressor in biota	Would include evidence of a chemical residue within an organism of interest	This criterion is similar to the US EPA's complete exposure pathway (Suter et al. 2002). However, we have generalized its meaning to include nonecotoxicological studies. Within our framework, this evidence should be reported but is not used in the weighting process. If present, this evidence may provide further support and strengthen confidence in the conclusion.
Agreement among hypotheses	When the results for the individual cause-effect hypotheses are considered collectively, do they support or refute the overall question (hypothesis) developed at Stage 1?	The final conclusion is always a matter of judgment by the reviewer. It requires collective consideration of the conclusions for each hypothesized cause-effect relationship. This metacriterion is applied at a different level in the analysis compared to the other criteria (i.e., across all hypotheses) and is most similar to the US EPA's considerations based on multiple lines of evidence (Suter et al. 2002).

^a Dose-response is a subset of evidence of response. In summing study weights to assess support for a hypothesis, the study-weight for a dose-response evidence item also will contribute to the summed study weight for evidence of response.

all are used to assess the weighted evidence (Table 4). The 1st criterion, *plausibility*, is addressed by the conceptual model (Step 3), which requires the reviewer to describe what he or she thinks are plausible cause-effect linkages. The next 3 criteria are assessed quantitatively using sums of the study weights. For *evidence of response*, the reviewer sums the study weights of all evidence items (i.e., from multiple studies) in favor of the hypothesis. Similarly, for *dose-response*, the reviewer sums the study weights of all evidence items that show a dose-response relationship in favor of the hypothesis. For *consistency of association*, the reviewer sums the study weights of

all evidence items that *do not* support the hypothesis (i.e., no evidence of a response even though the hypothesized cause was present, or an observed response in the direction opposite that hypothesized). The other 2 causal criteria, *evidence of stressor in biota* and *agreement among hypotheses* (Table 4), are used in the Eco Evidence method but not in the process of weighting the evidence.

When assessing the evidence for or against a causal relationship, Eco Evidence applies a threshold of 20 summed points for each of the weighted criteria above. Again, the threshold was derived after numerous trials and extensive consultation with

TABLE 5. Possible outcomes of an Eco Evidence analysis, using the default 20-point threshold of summed evidence weights. Note that the summed study-weight for dose–response does not affect the conclusion because it is, by definition, a subset of the summed evidence weight for evidence of response.

Summed evidence weights			Conclusion
Evidence of response	Dose–response	Consistency of association	
≥20	≥20	<20	Support for hypothesis
≥20	<20	<20	Support for hypothesis
<20	<20	<20	Insufficient evidence
≥20	≥20	≥20	Inconsistent evidence
≥20	<20	≥20	Inconsistent evidence
<20	<20	≥20	Support for alternative hypothesis

ecologists (discussed later in the paper). Like the study weight values (Table 3), the threshold can be altered as long as this alteration is documented and justified. The default 20-point threshold means that ≥3 independent, very-high-quality studies are sufficient to reach a conclusion concerning the presence (or absence) of a causal relationship. Conversely, ≥7 low-quality studies might be needed to reach the same conclusion. The threshold provides a convenient division of a continuous score in a way analogous to the almost-ubiquitous convention of 0.05 as a significant *p*-value, but, like significance levels, it should not be applied unthinkingly.

A range of outcomes is possible from the analysis, but a high level of support (i.e. ≥20 points for evidence of a response) and high consistency (i.e., <20 points for consistency of association) are both required to provide support for the cause–effect hypothesis (Table 5). The conclusion of *support for hypothesis* does not imply that the causal hypothesis has been proved. However, it is retained as a working hypothesis that may be falsified by future research (sensu Popper 1980). *Support for alternative hypothesis* is the falsification of the cause–effect hypothesis, and a new hypothesis should be sought. *Inconsistent evidence* is another form of falsification. In such a case, much evidence has been found *both for and against* the cause–effect hypothesis. In such circumstances, the reviewer should examine the scope of the hypothesis (Fig. 1). Investigation of the inconsistencies often reveals the reasons for a mixed response and provides the basis to refine the hypothesis (Steps 1–3). A refined review may reveal, for instance, the particular circumstances under which an environmental stressor and ecological response will be associated. The reviewer should document the findings from this re-examination and use them to restructure the literature review (Step 5). Last, *insufficient evidence* implies a knowledge gap in the literature and resulting opportunities for research.

Both the number of papers addressing the hypotheses and the consistency of their findings are important in providing confidence in the conclusions regarding causal relationships (Table 5). However, when summed evidence weights are close to the 20-point threshold, the reviewer must use care in interpreting the conclusions in Table 5. At this point, the reviewer also may draw support from the other causal criteria not so far applied in the framework: i.e., evidence of stressor in biota, and whether the response manifested as a dose–response relationship (Table 4).

Example conclusions: Harrison (2010) Question 1: Does macroinvertebrate community structure change as a result of accumulation of fine sediment within the stream bed?

Harrison (2010) found that published studies supported the hypotheses for a causal link between accumulation of fine sediment within the stream bed and decreased Ephemeroptera, Plecoptera, and Trichoptera (EPT taxa) richness, decreased Coleoptera abundance, and increased Oligochaeta abundance (Table 6). Evidence was inconsistent for change in total macroinvertebrate taxon richness, change in total macroinvertebrate abundance, decreased EPT abundance, and increased Chironomidae abundance (Table 6). Evidence was insufficient to assess whether accumulation of fine sediment causes a decrease in Coleoptera richness because no studies were found that were related to this hypothesis (scores of 0 in Table 6).

The final judgment

Last, the reviewer collectively considers the conclusions for the individual cause–effect hypotheses to determine the answer to the overall question (hypothesis) posed at Step 1 (*agreement among hypotheses* in Table 4). This stage is always a matter of considered judgement that depends on the nature of

TABLE 6. Numbers of citations and evidence items, and summed evidence weights and conclusions for the hypotheses within question 1: Does macroinvertebrate community structure change as a result of fine-sediment accumulation within the streambed? Values for conclusion are based on the rules in Table 5.

Effect (defined in conceptual model)	Number of citations	Number of evidence items		Summed evidence weights			Conclusion
		In favor of hypothesis	Not in favor of hypothesis	Evidence of response	Dose-response	Consistency of association	
Change in total macroinvertebrate taxon richness	27	23	7	151	115	53	Inconsistent
Change in total macroinvertebrate abundance	28	25	7	172	116	57	Inconsistent
Decreased EPT taxon richness	13	14	0	82	51	6	Support
Decreased EPT abundance	29	29	4	182	104	30	Inconsistent
Decreased Coleoptera richness	0	0	0	0	0	0	Insufficient
Decreased Coleoptera abundance	4	3	1	25	21	7	Support
Increased Oligochaeta abundance	13	11	2	86	42	13	Support
Increased Chironomidae abundance	21	9	13	65	28	59	Inconsistent

the original question. In the case study, conclusions of detailed cause–effect hypotheses relating to question 1 varied among individual taxonomic groups (Table 6) and also among the different quantifiable causal agents (results for questions 2 and 3 reported in Harrison 2010). However, overall sufficient evidence existed in the published ecological literature to conclude that addition of fine sediment to streams causes changes in macroinvertebrate community structure. Thus, the Eco Evidence process laid a strong foundation for a research project designed to address identified gaps in knowledge of the relationships between changes in macroinvertebrate community structure and fine sediments in rivers (Harrison 2010).

The Eco Evidence software provides a standard report at the end of any assessment. This report contains all the information used to undertake the assessment. It also contains any important caveats on the conclusions, including the particular conditions under which a particular stressor is associated with an environmental effect (refined hypothesis at Step 6), and important covariates that also may have causal relations with the response observed (multiple supported hypotheses at Step 7). The report clearly shows how the information was used to reach the overall conclusion, maximizing transparency and repeatability of the assessment and the conclusions drawn from the evidence.

Discussion

When does the lore surrounding the ecological effects of a given environmental stressor become law, ecologically speaking? Traditional approaches for inferring causality rely on rigorous experiments that include controls, randomized treatments, and sufficient replicates to provide the experimental power to draw clear conclusions (Downes et al. 2002, Johnson 2002). However, such experiments rarely can be performed in environmental studies, and individual investigations seldom achieve a severe test of the hypothesis. Faced with similar difficulties, epidemiologists developed the criteria of causation, and the causal inference approach has been adopted as a standard epidemiological tool.

The Eco Evidence analysis method has been designed for use in environmental sciences by adapting and developing the epidemiological approach. It provides a pragmatic framework for transparent and repeatable synthesis of evidence from the literature to assess support for and against questions of cause and effect, which may be site-specific or general. However, like the Popperian

paradigm of scientific progress, the method cannot ever provide proof of causal relations. Potential uses for Eco Evidence include: 1) to assess likely environmental impacts of a proposed development; 2) to identify the most likely cause(s) of an observed environmental impact; 3) to use the existing scientific literature to assess the more general applicability, or transferability, of results from local field studies; 4) to support transparent and defensible decision-making in environmental management and evidence-based planning; 5) to complement environmental risk assessments; 6) to present evidence in a transparent, repeatable, and defensible format suitable for use in legal cases or administrative action over environmental impacts; 7) to provide quality assurance for any literature review undertaken by consultants to environmental management organizations; and 8) to focus a literature review to the point where the output can be published as a succinct review paper (e.g., Harrison 2010, Greet et al. 2011).

Causal criteria in relation to induction, deduction, conjecture, and refutation

Much has been written concerning how the use of causal criteria fits into the 2 major modes of scientific logic—induction and deduction (Buck 1975, Susser 1986b, Morabia 1991, Beyers 1998, Ward 2009). Largely through the work of Karl Popper (1980), most scientific study conforms to the *conjecture–refutation* model, whereby hypotheses are tested and can be *falsified* (i.e., rejected) but never proved. The introduction of such concepts to the largely inductive field of epidemiology (Buck 1975) caused considerable controversy concerning those causal criteria deemed to be inductively driven (Morabia 1991). However, most scientists use a blend of inductive and deductive thinking to make scientific progress (Susser 1986b), and many hypotheses tested under the conjecture–refutation model are derived inductively (Beyers 1998).

With the above in mind, we largely retained the conjecture–refutation model familiar to most scientists in developing the Eco Evidence framework but paid little attention to whether individual aspects of the process were purely inductive, deductive, or neither. The reviewer uses evidence from the literature to test one or more cause–effect hypotheses, which may be developed either inductively or deductively. A clear finding of support for alternate hypothesis or inconsistent evidence should be regarded as falsification of the hypothesis, meaning that a new hypothesis should be sought (Rothman 2002). Conversely, a finding of support for hypothesis should be regarded as

corroboration of the causal hypothesis (sensu Popper 1980), rather than confirmation or proof. The hypothesis survives the test but is still unproved and may be falsified by future research. Indeed, in Eco Evidence, use of the 20-point threshold of summed study weights makes a falsification of nonrobust hypotheses almost inevitable (Harrison 2010). As more and more evidence is considered, very general cause–effect hypotheses that were previously supported are likely to be falsified by a finding of inconsistent evidence because of varying responses in studies that considered different levels of the stressor, locations, environment types, species, etc. In such a case, a more specific hypothesis is less likely to be falsified. Thus, the 8-step framework presented here leads naturally to generation of more detailed cause–effect hypotheses as more evidence becomes available—a process similar to that by which science advances in general.

The limits of Eco Evidence and how it compares other synthesis techniques

Causal inference techniques, such as Eco Evidence, cannot prove causality (Fox 1991, Suter et al. 2010). In that regard, Eco Evidence is no different from any statistical analysis of observational data (Greenland 1998). What Eco Evidence does provide is a rigorous assessment of the evidence for and against cause–effect questions.

Eco Evidence may be used to assess questions of either general or specific causation. Results may be used to support a conclusion of causation relating to a site-specific association of an environmental stressor and observed impairment. Alternatively, they may be used to reach generalized conclusions in a systematic review. Such general conclusions are important for assessing the state of knowledge in a research area and can be used to test (and occasionally debunk) commonly held assumptions in ecology (e.g., Stewart et al. 2006) or to identify knowledge gaps requiring further research (e.g., Greet et al. 2011). The US EPA's CADDIS method (Norton et al. 2008) is another causal assessment framework for synthesizing evidence from several sources. CADDIS concentrates mostly on questions of specific causation because it was developed for identifying site-specific causes of environmental impairment. CADDIS also differs specifically from Eco Evidence in that it does not weight evidence according to strength of the study, is more focused on inclusion of data sets rather than published studies, and does not actually require use of the literature in an assessment. CADDIS is not supported by analysis software analogous to Eco Evidence, but the EPA has created a number of

downloadable aids to assist an assessment (available from: www.epa.gov/caddis). Meta-analysis is another cross-study synthesis technique that may be familiar to most ecologists (Osenberg et al. 1999). Causal criteria analysis is not meta-analysis, but it can provide a complementary, rather than an alternative, technique. Meta-analysis is concerned with estimating an ensemble effect size across a number of studies (Gurevitch and Hedges 2001, Sutton and Higgins 2008). No requirement exists in meta-analysis to assess the causal plausibility of the association under investigation, and Weed (2000) argues that meta-analysis should not be used to reach causal conclusions. However, the robust statistical approaches used in meta-analysis potentially could be used to precisely quantify strength of association and consistency of association in a combined analysis (Weed 2000). Such an approach may not be well suited to environmental sciences, where a large proportion of studies fail to report the summary statistics necessary for inclusion in a meta-analysis (Greet et al. 2011). The Eco Evidence approach, as presented here, allows the reviewer to include a greater range of literature in their analysis (Greet et al. 2011).

All review techniques are affected by publication bias—the tendency to publish only significant results (Koricheva 2003). Some analytic techniques to account for publication bias exist for explicitly numerical synthesis methods like meta-analysis (e.g., funnel plots; Song et al. 2002), but such techniques cannot be applied to Eco Evidence. However, we think that publication bias may be less of an issue for Eco Evidence than it is for techniques such as meta-analysis. Eco Evidence can include evidence even when the summary statistics necessary for meta-analysis are not supplied. Authors often report particular factors in analyses as being nonsignificant without summary statistics, but then provide full summary statistics for significant results. In an Eco Evidence analysis, both results could be included, whereas only the significant result could be included in a meta-analysis.

Eco Evidence uses systematic review of the literature as a relatively underused resource for increasing inferential power (Downes et al. 2002). Systematic review techniques allow researchers to produce succinct reviews that test clearly stated hypotheses. This type of review is in contrast to the longer narrative reviews more familiar to environmental scientists. Narrative reviews tend to provide a comprehensive coverage of the literature, but seldom provide any assessment of the relative strength of evidence for or against cause–effect hypotheses. In the example herein, different conclusions were reached

for different specific hypotheses because the relative strength of evidence differed between them. We cannot be confident that a narrative review would reach the same conclusion because such reviews have no rules for interpreting evidence. Systematic review is a common tool used particularly in the health sciences. Initiatives like the Cochrane Collaboration (<http://www.cochrane.org>) have used systematic reviews to drive an ‘effectiveness revolution’ via incorporation of scientific evidence into clinical practice (Stevens and Milne 1997). Systematic reviews are underused in environmental sciences (Pullin and Stewart 2006, Pullin et al. 2009). Interest in using systematic review to guide environmental management decisions is increasing (Pullin and Knight 2001, Sutherland et al. 2004), but little guidance exists on appropriate methods to use or the tools to implement them. The Eco Evidence analysis method and associated software helps fill this gap.

Strengths of the Eco Evidence method

Despite calls for use of causal criteria in environmental science (Fox 1991, Downes et al. 2002, Suter et al. 2002, Adams 2005, Plowright et al. 2008), these methods have not been widely adopted, and existing case studies lack consistency and clarity in the criteria used. Both of these observations could be explained by the lack of well developed methods for applying the criteria. The Eco Evidence method provides a standardized framework within which environmental scientists can conduct rigorous causal criteria analyses. Some authors have argued against standardizing causal-inference methods (Downes et al. 2002), but others have argued that greater standardization of definitions and greater specification of methods for synthesizing across criteria would be beneficial to practitioners (Weed 1997, 2002). Advantages to using a standardized approach include the transparency and repeatability of analyses, both of which are important for environmental management decisions subject to legal challenge or exposed to political scrutiny (Norris et al. 2005). Several closely related attempts have been made to score the contribution of individual criteria in arguing a case for causation (Susser 1986a, Fox 1991, Suter et al. 2010). However, a numeric method for combining these scores to aid an overall judgment was not used in these attempts. Eco Evidence does provide such a method and has the added beneficial capability of being able to identify knowledge gaps where insufficient evidence exists to reach a conclusion concerning a cause–effect hypothesis. This conclusion is more useful than a failure to reject a null hypothesis and guards against type-II errors

(i.e., false negatives) by specifying minimum evidence levels required to reach a conclusion. Such a conclusion also illustrates where empirical research may be needed, and thus, is useful for research planning.

A major advance in Eco Evidence over other causal-analysis methods is the weighting of individual studies in the overall assessment according to the strength of the study design. In investigations of existing literature, epidemiologists tend to disregard all studies that do not conform to the gold standard of a randomized-control trial and examine results from the less-powerful cohort and case-control studies only when randomized-control trial data are unavailable (Tugwell and Haynes 2006). Similarly, Best Evidence Synthesis (Slavin 1995) advocates systematic analysis of only a subset of studies that have the single best design. In environmental science, sufficient studies on a particular question seldom will exist to allow a reviewer to focus only on those of the highest quality. Instead, we must make use of all the available evidence. The weighting scheme in Eco Evidence allows use of all published evidence that relates to a particular question, including both observational and experimental studies, but ensures that results of stronger studies—those less likely to be affected by confounding—play a relatively larger part in the assessment of causality than results from weaker studies. This feature is a major step forward and has potential application for causal inference in other discipline areas where empirical research may be scant.

Collecting evidence from the literature is time consuming. One estimate is that each relevant study takes ~1 h to process (Webb et al., in press a). However, in systematic literature reviews, studies and the evidence they yield are data. The time taken to extract evidence from a published paper compares favorably to the time needed to collect data in the laboratory or field. Moreover, the Eco Evidence software includes an online database (Wealands et al. 2009, Webb et al., in press b) developed specifically so that evidence would be available for reuse in future research, thereby reducing overall effort compared to other systematic review techniques.

Other comments

In designing the Eco Evidence framework, we considered whether the causal criteria commonly used in epidemiology (Hill 1965, Weed and Gorelic 1996, Tugwell and Haynes 2006) could be practically applied to environmental questions. The current framework does not make use of the strength of

association criterion, but this is one of the most common criteria used in epidemiology (Weed and Gorelic 1996, Tugwell and Haynes 2006). Strength of association in epidemiology is usually expressed as relative risk (Susser 1986a), but that measure does not translate well to the continuous measures of effect commonly encountered in ecology (but see Van Sickle et al. 2006, Van Sickle and Paulsen 2008). Fox (1991) and Fabricius and De'Ath (2004) have argued that, for ecology, measured ecological effect size is equally valid as a measure of strength of association. The database in the Eco Evidence software (eWater CRC 2010) currently has fields for capturing information on effect size. A future version of the Eco Evidence algorithm may use this information as part of the formal assessment of cause–effect hypotheses, but we have to develop robust methods for combining such evidence with evidence on consistency of association. Such a development is not straightforward because these different types of evidence do not exist on a common scale (Suter and Cormier 2011).

Other potential revisions for future versions of the algorithm include more detailed classifications of study designs. The framework currently does not distinguish between observational studies and true experiments (i.e., studies in which treatments were randomized). However, experiments provide more powerful arguments for causality. Other revisions could reduce the likelihood of confounding. For instance, gradient designs might be assigned a lower score unless study authors demonstrate they have accounted for confounding influences. Similarly, in assessing consistency of association, greater weight could be placed on combinations of consistent studies from different environments and that used different techniques (Hill 1965). Such sets of studies are less likely to suffer from unidentified confounding, although the possibility can never be eliminated entirely. Another possibility is to weight some evidence as highly relevant because of its close association with the type of system under study (e.g., area, source of stress). For example, when assessing causation at a specific location, a reviewer could weight evidence from that site more highly than evidence from other sites. Such an approach would align Eco Evidence more closely with the CADDIS framework (Norton et al. 2008). However, these revisions would complicate what is currently an appealingly simple scoring system. Such extra complication should be considered only if it were to improve substantially our confidence in the resulting conclusions.

The choice of values for study weights and the 20-point threshold as defaults in Eco Evidence have been

criticized for the seemingly arbitrariness of the values. Nevertheless, we think they are useful values, particularly because they were developed through expert consultation with ecologists and have produced intuitively sensible results in case studies to date. The expert-consultation process explored the relative ability of different study designs to identify causal relationships by reducing the likelihood of confounding and how replication (as opposed to study design per se) strengthened this ability. The number of study design types and bands of replication were kept to a workable minimum to maximize reproducibility of results between different reviewers (Norris et al. 2005). Reproducibility will be subject to formal assessment in future, but some opportunistically gathered data from 2 case studies appears to indicate high reproducibility (Webb et al., in press a). The threshold was developed by exploring the number of consistent results from high- or low-quality studies that experts needed to see to have a high level of confidence in the existence (or otherwise) of a causal relation. The reviewer can also redefine both weights and thresholds as long as the reason for such a change is documented to maintain transparency. Moreover, the final judgment at the end of Step 8 requires an intelligent synthesis of the results. Judgment sits apart from the algorithmic component of Eco Evidence.

Conclusion

Considerable impetus exists to move from experience-based to evidence-based methods in environmental sciences, particularly when answering those questions relevant to management decisions (Pullin and Knight 2001, Sutherland et al. 2004). The 8-step Eco Evidence method, supported by the Eco Evidence software and the evidence report it produces, provides a transparent and repeatable framework for assessing the evidence for and against causal relationships in terrestrial and aquatic environments. Assessments conducted using this method can inform decision-making for environmental management, potentially leading to improved outcomes for all stakeholders. We do not consider the method presented here to be the last word on causal inference in environmental sciences. Rather, we expect the method to evolve and to improve with rigorous testing across a range of different subject areas. Provision of a framework for quantifying and combining the evidence, along with the major advance of a weighting system for individual studies, is an important first step in facilitating broader use by the research community of systematic methods to assess cause-effect relationships in environmental sciences.

Acknowledgements

We wish to acknowledge and thank the researchers involved in the early development of the Eco Evidence method and software (Peter Liston, James Mugodo, Gerry Quinn, Peter Cottingham, Leon Metzeling, Stephen Perris, David Robinson, David Tiller, Glen Wilson, Gail Ransom, and Sam Silva), Steve Wealands and Patrick Lea for developing the released version of the Eco Evidence software, and the many eWater staff and students who were involved in product testing. This manuscript benefited from the careful review of Sue Norton, Glen Suter, and an anonymous referee, and from editorial review by Ann Milligan. The development of Eco Evidence has been funded by the eWater Cooperative Research Centre and the Cooperative Research Centre for Freshwater Ecology.

Literature Cited

- ADAMS, S. M. 2005. Assessing cause and effect of multiple stressors on marine systems. *Marine Pollution Bulletin* 51:649–657.
- BEYERS, D. W. 1998. Causal inference in environmental impact studies. *Journal of the North American Benthological Society* 17:367–373.
- BUCK, C. 1975. Popper's philosophy for epidemiologists. *International Journal of Epidemiology* 4:159–168.
- BURKHARDT-HOLM, P., AND K. SCHEURER. 2007. Application of the weight-of-evidence approach to assess the decline of brown trout (*Salmo trutta*) in Swiss rivers. *Aquatic Sciences* 69:51–70.
- COLLIER, T. K. 2003. Forensic ecotoxicology: establishing causality between contaminants and biological effects in field studies. *Human and Ecological Risk Assessment* 9: 259–266.
- CORMIER, S. M., S. B. NORTON, G. W. SUTER, D. ALTFATER, AND B. COUNTS. 2002. Determining the causes of impairments in the Little Scioto River, Ohio, USA: part 2. Characterization of causes. *Environmental Toxicology and Chemistry* 21:1125–1137.
- CORMIER, S. M., G. W. SUTER, AND S. B. NORTON. 2010. Causal characteristics for ecoepidemiology. *Human and Ecological Risk Assessment* 16:53–73.
- DOWNES, B. J., L. A. BARMUTA, P. G. FAIRWEATHER, D. P. FAITH, M. J. KEOUGH, P. S. LAKE, B. D. MAPSTONE, AND G. P. QUINN. 2002. *Monitoring ecological impacts: concepts and practice in flowing waters*. Cambridge University Press, Cambridge, UK.
- EVANS, A. S. 1976. Causation and disease: Henle-Koch postulates revisited. *Yale Journal of Biology and Medicine* 49:175–195.
- EWater CRC (eWater Cooperative Research Centre). 2010. *Eco Evidence: a systematic approach to evaluate causality in environmental science*. eWater Cooperative Research Centre, Canberra, Australia. (Available from: www.toolkit.net.au/tools/eco-evidence)

- FABRICIUS, K. E., AND G. DE'ATH. 2004. Identifying ecological change and its causes: a case study on coral reefs. *Ecological Applications* 14:1448–1465.
- FOX, G. A. 1991. Practical causal inference for ecoepidemiologists. *Journal of Toxicology and Environmental Health* 33:359–373.
- GREEN, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York.
- GREENLAND, S. 1998. Meta-analysis. Pages 643–673 in K. J. Rothman and S. Greenland (editors). *Modern epidemiology*. Lippincott-Raven, Philadelphia, Pennsylvania.
- GREET, J., J. A. WEBB, AND R. D. COUSENS. 2011. The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis. *Freshwater Biology* 56:1231–1247.
- GROVE, J. R., J. A. WEBB, P. M. MARREN, M. J. STEWARDSON, AND S. R. WEALANDS. High and dry: an investigation using the causal criteria methodology to investigate the effects of regulation, and subsequent environmental flows, on floodplain geomorphology. *Wetlands* (in press). doi: 10.1007/s13157-011-0253-9
- GUREVITCH, J., AND L. V. HEDGES. 2001. Meta-analysis: combining the results of independent experiments. Pages 346–369 in S. M. Scheiner and J. Gurevitch (editors). *Design and analysis of ecological experiments*. Oxford University Press, New York.
- HAAKE, D. M., T. WILTON, K. KRIER, A. J. STEWART, AND S. M. CORMIER. 2010. Causal assessment of biological impairment in the Little Floyd River, Iowa, USA. *Human and Ecological Risk Assessment* 16:116–148.
- HARRISON, E. T. 2010. Fine sediment in rivers: scale of ecological outcomes. PhD Dissertation, University of Canberra, Canberra, Australia. (Available from: <http://tinyurl.com/Harrison-2010>)
- HILL, A. B. 1965. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 58:295–300.
- JOHNSON, D. H. 2002. The importance of replication in wildlife research. *Journal of Wildlife Management* 66: 919–932.
- KORICHEVA, J. 2003. Non-significant results in ecology: a burden or a blessing in disguise? *Oikos* 102:397–401.
- LOWELL, R. B., J. M. CULP, AND M. G. DUBE. 2000. A weight-of-evidence approach for Northern River risk assessment: integrating the effects of multiple stressors. *Environmental Toxicology and Chemistry* 19:1182–1190.
- MORABIA, A. 1991. On the origin of Hill's causal criteria. *Epidemiology* 2:367–369.
- NICHOLS, S., A. WEBB, R. NORRIS, AND M. STEWARDSON. 2011. *Eco Evidence analysis methods manual: a systematic approach to evaluate causality in environmental science*. eWater Cooperative Research Centre, Canberra, Australia. (Available from: <http://tinyurl.com/Eco-Evidence-manual>)
- NORRIS, R. H., P. LISTON, J. MUGODO, S. NICHOLS, G. P. QUINN, P. COTTINGHAM, L. METZELING, S. PERRISS, D. ROBINSON, D. TILLER, AND G. WILSON. 2005. Multiple lines and levels of evidence for detecting ecological responses to management intervention. Pages 456–463 in I. D. Rutherford, I. Wiszniewski, M. J. Askey-Doran, and R. Glazik (editors). *Proceedings of the 4th Australian Stream Management Conference: Linking Rivers to Landscapes*. Department of Primary Industries, Water and Environment, Launceston, Tasmania. (Available from: <http://tinyurl.com/Norris-et-al-2005>)
- NORTON, S. B., S. M. CORMIER, G. W. SUTER, K. SCHOFIELD, L. YUAN, P. SHAW-ALLEN, AND C. R. ZIEGLER. 2008. CADDIS: the causal analysis/diagnosis decision information system. Pages 351–374 in A. Marcomini, G. W. Suter, and A. Critto (editors). *Decision support systems for risk-based management of contaminated sites*. Springer, New York.
- OSENBERG, C. W., O. SARNELLE, AND D. E. GOLDBERG. 1999. Meta-analysis in ecology: concepts, statistics, and applications. *Ecology* 80:1103–1104.
- PLATT, J. R. 1964. Strong inference: certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* 146:347–353.
- PLOWRIGHT, R. K., S. H. SOKOLOW, M. E. GORMAN, P. DASZAK, AND J. E. FOLEY. 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment* 6:420–429.
- POPPER, K. R. 1980. *The logic of scientific discovery*. Hutchinson, London, UK.
- POPPER, K. R. 1983. *Realism and the aim of science: postscript to the logic of scientific discovery*. Hutchinson, London, UK.
- PULLIN, A. S., AND T. M. KNIGHT. 2001. Effectiveness in conservation practice: pointers from medicine and public health. *Conservation Biology* 15:50–54.
- PULLIN, A. S., T. M. KNIGHT, AND A. R. WATKINSON. 2009. Linking reductionist science and holistic policy using systematic reviews: unpacking environmental policy questions to construct an evidence-based framework. *Journal of Applied Ecology* 46:970–975.
- PULLIN, A. S., AND G. B. STEWART. 2006. Guidelines for systematic review in conservation and environmental management. *Conservation Biology* 20:1647–1656.
- ROTHMAN, K. J. 2002. *Epidemiology: an introduction*. Oxford University Press, New York.
- ROTHMAN, K. J., S. GREENLAND, AND T. L. LASH. 2008. *Modern epidemiology*. Wolters Kluwer Health/Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.
- SLAVIN, R. E. 1995. Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology* 48:9–18.
- SONG, F. J., K. S. KHAN, J. DINNES, AND A. J. SUTTON. 2002. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology* 31:88–95.
- STEVENS, A., AND R. MILNE. 1997. The effectiveness revolution and public health. Pages 197–225 in G. Scally (editor). *Progress in public health*. Royal Society of Medicine Press, London, UK.
- STEWART, G. B., H. R. BAYLISS, D. A. SHOWLER, A. S. PULLIN, AND W. J. SUTHERLAND. 2006. Does the use of in-stream structures and woody debris increase the abundance of

- salmonids? Collaboration for Environmental Evidence. (Available from: <http://www.environmentalevidence.org/SR12.html>)
- SUSSER, M. 1986a. Rules of inference in epidemiology. *Regulatory Toxicology and Pharmacology* 6:116–128.
- SUSSER, M. 1986b. The logic of Sir Karl Popper and the practice of epidemiology. *American Journal of Epidemiology* 124:711–718.
- SUTER, G. W., AND S. M. CORMIER. 2011. Why and how to combine evidence in environmental assessments: weighing evidence and building cases. *Science of the Total Environment* 409:1406–1417.
- SUTER, G. W., S. B. NORTON, AND S. M. CORMIER. 2002. A methodology for inferring the causes of observed impairments in aquatic ecosystems. *Environmental Toxicology and Chemistry* 21:1101–1111.
- SUTER, G. W., S. B. NORTON, AND S. M. CORMIER. 2010. The science and philosophy of a method for assessing environmental causes. *Human and Ecological Risk Assessment* 16:19–34.
- SUTHERLAND, W. J., A. S. PULLIN, P. M. DOLMAN, AND T. M. KNIGHT. 2004. The need for evidence-based conservation. *Trends in Ecology and Evolution* 19:305–308.
- SUTTON, A. J., AND J. P. I. HIGGINS. 2008. Recent developments in meta-analysis. *Statistics in Medicine* 27:625–650.
- TUGWELL, B., AND R. B. HAYNES. 2006. Assessing claims of causation. Pages 356–387 in R. B. Haynes, D. L. Sackett, G. H. Guyatt, and P. Tugwell (editors). *Clinical epidemiology: how to do clinical practice research*. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.
- UNDERWOOD, A. J. 1997. *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, New York.
- USDHEW (US DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE). 1964. Smoking and health. Report of the advisory committee to the Surgeon General of the Public Health Service. US Department of Health, Education, and Welfare, Washington, DC.
- VAN SICKLE, J., AND S. G. PAULSEN. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. *Journal of the North American Benthological Society* 27:920–931.
- VAN SICKLE, J., J. L. STODDARD, S. G. PAULSEN, AND A. R. OLSEN. 2006. Using relative risk to compare the effects of aquatic stressors at a regional scale. *Environmental Management* 38:1020–1030.
- WARD, A. C. 2009. The role of causal criteria in causal inferences: Bradford Hill's "aspects of association". *Epidemiologic Perspectives and Innovations* 6. doi:10.1186/1742-5573-6-2
- WEALANDS, S. R., J. A. WEBB, AND M. J. STEWARDSON. 2009. Evidence-based model structure: the role of causal analysis in hydro-ecological modelling. Pages 2465–2471 in R. S. Anderssen, R. D. Braddock, and L. T. H. Newham (editors). 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Australia. (Available from: <http://tinyurl.com/ybcb3we>)
- WEBB, J. A., S. J. NICHOLS, R. H. NORRIS, M. J. STEWARDSON, S. R. WEALANDS, AND P. LEA. Ecological responses to flow alteration: assessing causal relationships with Eco Evidence. *Wetlands* (in press a). doi: 10.1007/s13157-011-0249-5
- WEBB, J. A., S. R. WEALANDS, P. LEA, S. J. NICHOLS, S. C. DE LITTLE, M. J. STEWARDSON, AND R. H. NORRIS. Eco Evidence: using the scientific literature to inform evidence-based decision making in environmental management. In F. Chan, D. Marinova and R. S. Anderssen (editors). MODSIM2011 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Australia (in press b).
- WEED, D. L. 1997. On the use of causal criteria. *International Journal of Epidemiology* 26:1137–1141.
- WEED, D. L. 2000. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *International Journal of Epidemiology* 29:387–390.
- WEED, D. L. 2002. Environmental epidemiology: basics and proof of cause–effect. *Toxicology* 181:399–403.
- WEED, D. L., AND L. S. GORELIC. 1996. The practice of causal inference in cancer epidemiology. *Cancer Epidemiology Biomarkers and Prevention* 5:303–311.
- WISEMAN, C. D., M. LEMOINE, AND S. CORMIER. 2010. Assessment of probable causes of reduced aquatic life in the Touchet River, Washington, USA. *Human and Ecological Risk Assessment* 16:87–115.

Received: 24 March 2011

Accepted: 14 September 2011