Causal Criteria Methods Manual

Methods for applying the multiple lines and levels of evidence (MLLE) approach for addressing questions of causality

Richard Norris, Susan Nichols, Gail Ransom, Peter Liston, Amara Barlow, and James Mugodo

eWater Cooperative Research Centre, &

Institute for Applied Ecology
School of Resource Environment & Heritage Science, University of Canberra.

Revised May 2008

UNIVERSITY OF CANBERRA
Institute for Applied Ecology
Ecological Solutions for a Healthy Environment

Introduction

This manual describes how a Multiple Lines and Levels of Evidence (MLLE) approach is used to develop 'Causal Criteria' to address questions of causality between environmental stressors, management interventions and ecological outcomes (Norris et al. 2005). The approach is also applicable to any technical literature that aims to demonstrate a relationship between an apparent cause and an apparent effect.

This approach investigates various lines of evidence, these are system attributes (ecosystem or other) e.g. tadpole abundance, macroinvertebrate species richness, macrophyte biomass, number of fish abnormalities, water quality attributes, which are investigated in relation to a causal agent or stressor. Causal Criteria are then used to determine the case for inferring that a given agent causes a particular change in the system. The various types of evidence can collectively build a sufficiently strong case to infer causality.

An 8-step framework provides a method for drawing together information from different data sources and the scientific literature to strengthen conclusions on how a particular causal agent, stressor, human activity or natural event, may influence the environment or system of interest. Scientists and managers are commonly faced with a situation where the information from the various sources provides conflicting results or advice. Therefore, it is important to have a transparent, consistent and logical framework to evaluate all the information and provide confidence for a strong conclusion.

The MLLE approach was originally developed for studies in epidemiology (medical science), where a lack of experimental data resulted in a weak ability to draw inferences about causality. Many ecological, hydrological, or other types of studies may also have limited opportunity for proper replication and randomization of treatments, thus weakening our ability to draw inferences. The ‘Causal Criteria’ method applies the MLLE approach to strengthen our inferential ability.

The Causal Criteria procedure involves reviewing existing literature to summarize and synthesize relevant research on a topic. The quality of a literature review is often dependent on elements that introduce subjectivity and bias, such as:

- the thoroughness of the writer's search
- the quality and reliability of the writer's sources
- the ability of the writer to relate research studies to one another and to the writer's own purpose; and
- the objectivity of the writer in selecting, interpreting, organizing, and summarizing the research he or she has reviewed

The Causal Criteria approach can reduce some of the individual’s bias and subjectivity, can provide an indication on whether sufficient evidence has been collected and focuses on the quality of studies reviewed. The quality of the evidence (whether from the existing

literature or your own unpublished study) is evaluated in terms of three study quality attributes:

- study design type
- number of independent sampling units used as control (e.g. reference sites)and
- number of [potentially] impacted independent sampling units (e.g. test sites, treatment locations).

The results section of any given study may report on more than one effect, e.g. fish abundance, fish taxa richness and a macroinvertebrate index, or salinity values and water flow measurements. For each of these lines of evidence the study quality attributes may differ if they have been investigated with different levels of rigour. Also, the evidence within and among different scientific papers can have different reasons for being relevant (or not) to the area of under investigation. The Causal Criteria method requires reviewers to explicitly record the relevance, or lack of relevance, of each line of evidence (i.e. effect) in a given article.

A software program is available to assist with the application of the Causal Criteria framework to a research question. It comprises a customized database, which requires the user to enter details from the literature review (including references) in a standardized and overt fashion. The software program allows the user to consistently store information and efficiently access the information, and assists in weighting the information in relation to its quality. The program also summarizes the evidence by generating a report that evaluates the strength of support for your hypothesis using the overt set of rules provided by the Causal Criteria method.

This Causal Criteria methods manual provides information about the 8 steps of the Causal Criteria framework; looks specifically at the process for weighting papers; and provides guidelines on how to interpret the outputs. An example, of a Causal Criteria application is demonstrated in the Appendices section.

The 8-Step Causal Criteria Framework

The 8-step Causal Criteria framework (Fig. 1) is modified from Downes et al. (2002) and is used to assess the support for, and quality of, cause and effect evidence. The first steps are designed to develop the specific question to which the Causal Criteria framework will be applied and will put the question into context.

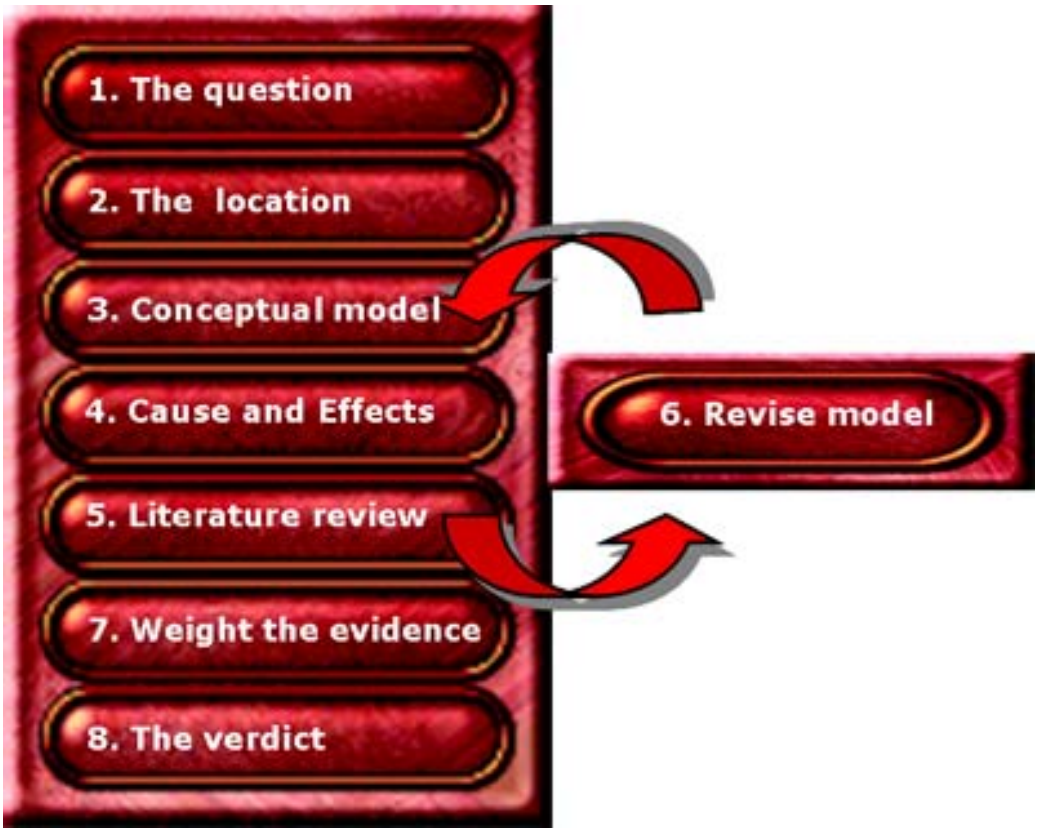


Figure1. Steps in applying the Causal Criteria framework

Note that we use of the term “causal agent” rather than “stressor” in the documentation. Causal agent may be a better word to use because the agents of cause are not always stressors, e.g. flows for the environment.

Step 1

Document the nature of the problem and draft the question under investigation

The first step is to document the nature of the problem under investigation, the potential agent of cause, the human activity, etc. and consider the potential impact, including the timing, size and likely magnitudes of any effects.

As an example for ecological studies, the Cotter River has been dammed to create a water supply for Canberra. Water trapped by Bendora Dam is diverted to Canberra for consumption. The flow regime below the Dam has been modified in a number of ways, principally reduction in mean annual flow and reduction in the frequency of small flows. The dam has a multi-level off-take and any downstream releases are generally of a similar temperature to inflow water. However, water quality will sometimes depend on the management of the dam for other problems e.g. release of layers of dirty water.

If more than one agent of cause is under investigation then you will need several questions and several assessments, i.e. each component should be a stand-alone question.

Thus:

- Has the reduction in mean annual flow damaged the aquatic ecosystem?
- Does reduction in the frequency of small flows damage the aquatic ecosystem?
- Do dirty-water releases damage the aquatic ecosystem?

Task:

- In this step you should document the nature of the problem under investigation, the potential causal agent, the human activity, potential effects, etc., and draft one or more questions for investigation.

Step 2

Document the characteristics of the investigation location

In what context will the question be asked?

The idea is to set the context and boundaries for the question under investigation.

It is necessary to decide whether the literature review will be restricted to studies with similar climatic regimes, geomorphology or other environmental features to the location under investigation. Therefore, in this step you describe the details of the investigation location.

For example, the Cotter River is an upland river (700-900m) in SE Australia with high aseasonal rainfall (1000mm), and a steep vegetated catchment. The river has a steep gradient and its substrate is predominantly cobble. It has a constrained channel with much bedrock outcropping. Where the river is impounded the catchment is a national park and has native vegetation. Downstream of the water supply impoundments land-uses are softwood forestry and the catchment also has areas of native vegetation. Public access to the water supply catchment area is restricted and so the principal potential impact is river regulation.

Task:

- Describe and document the details of your investigation location.

Step 3

Develop a conceptual model and clarify the question

Develop a conceptual model and identify the potential causal relationships. You may also need to refine the question you drafted in Step 1.

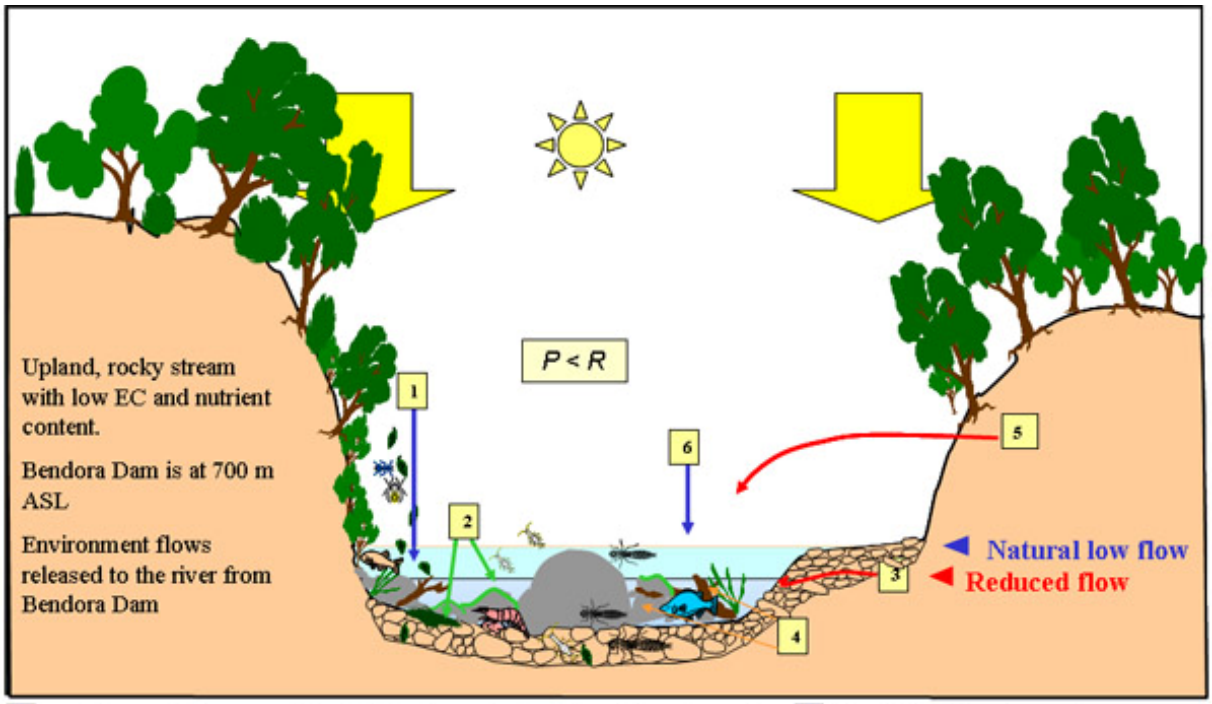
The operational question/hypothesis should articulate the "**quantifiable**" causal agent and the "**quantifiable**" effects.

Note that the framework is being used to address questions of causality.

For example, has the reduction in mean annual flow in the Cotter River, below Bendora Dam, resulted in a degradation of the aquatic ecosystem? The degradation of the ecosystem measured by the following effects (lines of evidence):

- Change in fish community;
- Change in algae;
- Change in riparian condition;
- Change in macrophytes;
- Decrease in macroinvertebrate taxa richness or diversity;
- Decrease in macroinvertebrate abundance or density; and
- Change in the predicted macroinvertebrate community composition.

The need for a good conceptual model is paramount, thus the following suggestions are provided, which provide some guidance on how to develop an appropriate conceptual model for the Causal Criteria application. An example of a conceptual diagram is shown below.



1 Riparian vegetation encroaches into the channel and reduces channel capacity. 2 Unpalatable filamentous algae accumulates. 3 Reduced flow results in armouring, reduced flushing of detritus, nutrients, fine sediment. 4 Habitat space for macroinvertebrates and fish in the substratum is reduced because of armouring and infilling with fine sediments. Also, some parts of the bottom may be exposed. 5 Sediment and organic matter may enter the channel directly from adjacent valley slopes and may not be flushed with low flows in the main channel. 6 Water quality dependant on releases from Bendora Dam (Canberra's water supply) – further downstream water quality (and quantity) also influenced by tributary and groundwater input.

Example of a conceptual model for the Cotter River below Bendora Dam, ACT

How to develop a conceptual model

1. Describe and define clearly the area of interest, and define the cause and/or effect of interest. Remember what prompted the investigation. Common reasons are the need to establish the cause for a biological effect (e.g. fish kill), or to identify the biological effects of a particular cause (e.g. change in flow). This information will help build the conceptual model and determine relevance of studies when reviewing the literature.

List important system descriptors e.g. upland river (700-900m) with many tributaries; in SE Australia; high aseasonal rainfall (1000mm); steep vegetated catchment; substrate is predominantly cobble; the impounded part of the catchment is a national park and has native vegetation.

List and describe the important components in the ecosystem

- Important geomorphologic components
- Major biological components
- Major processes operating and the relative importance of different pathways

2. Identify the cause or effect of interest in this schema and identify its relationship to other components / processes in the conceptual model. If possible, develop the model in consultation with others (such as other stakeholders or experts) to help ensure that all the relevant components of the system are identified. Care must be taken to ensure that the model captures the important relationships within the system, and does not become burdened by consideration of the minor relationships. An initial conceptual model may need to be pruned back to make it more parsimonious. These steps will help in the identification of appropriate lines of evidence (such as biological responses) and is also used to convey the investigator's system understanding to others.

- If the reason for the investigation is a cause (e.g. change in flow), this will have top down consequences. The conceptual model should be of a 'pristine' ecosystem so that the effects of the change of interest can be traced.
- If the reason for the investigation is an effect (e.g. fish kill), this could be the result of a range of yet to be identified causes. The conceptual model should be 'locally realistic', including all potential human influences so that the most likely can be identified in the conceptual model in order to clearly frame the question that the Causal Criteria will be used to address. Where multiple causal agents or stressors are potentially contributing to an effect, the aim is to identify the dominant causal agent i.e. that which makes the largest

contribution.

- List the candidate effects e.g. fewer invertebrate taxa than expected; change in diatom community composition.

3. Define the scope of the investigation and document temporal and spatial issues.

For example:

- The scope of the study needs to be limited (e.g. does a particular causal agent result in biological impairment?). When applying the Causal Criteria framework it is best to be specific with your question, rather than considering very broad issues, unless they can be considered as a single scenario.
- Consider temporal issues such as season. Will studies conducted in one season provide adequate evidence or do you need to consider only studies conducted over two or more seasons?
- If you are interested in a population-change response then consider generation times.
- Are you interested in the effects of a fire after 1 week or 10 years?
- At what distance downstream of the dam do you expect a response?

4. You are now ready to articulate an operational question.

- What is the "**quantifiable**" causal agent?
- What are the "**quantifiable**" effects?
- What is the appropriate resolution for the potential effects? For example, will family-level taxonomic resolution capture a homogeneous response of a particular guild?
- What are the spatial and temporal issues?

Task:

- Develop a conceptual model of the relationships in question, and be sure to identify the potential cause and effects, then refine your question.
- The operational question should articulate the "**quantifiable**" causal agent and the "**quantifiable**" effects.

Step 4

Decide on the relevant cause and effects, and list them

How will an effect be determined?

The previous step should have provided a very specific question for your investigation and you should be able to articulate the "quantifiable" causal agent and the "quantifiable" effects.

The effect: The level of resolution of an effect is important. For example, is the potential effect any change to a fish community, or is it change to the population size for one particular species?

There is a danger of losing too much information or inappropriately mixing up responses if we group the responses inappropriately. For instance, if you were investigating a change to fish community it may not always be appropriate to pool the species-richness response with the biomass response.

At first, divide lines of evidence into major categories e.g. fish, algae, macrophytes, macroinvertebrates etc. Below this level the definition of the potential effects would be guided by your conceptual model for the hypothesis being investigated. Using the conceptual model as a guide, it may be more appropriate to fit effects into broad lines of evidence, e.g. a change to the macroinvertebrate community. In other cases it would be best to use more specific effects as they would be described in the source literature or study under evaluation; i.e. an increase in or a decrease in macroinvertebrate biomass relative to 'normal'.

Note: Step 4 is an iterative step. Following the literature search the conceptual model and potential causal agents and effects should be reviewed.

Task:

- Document the relevant causal agents and potential effects for your current investigation.

Step 5

Conduct the literature search and review

Review the literature to locate the studies judged to be relevant to the investigation.

It is up to you to determine how many papers to include but a short literature review generally requires about 15 - 20 research articles.

Only two high quality studies (i.e. BACI study design, several control and impact locations) relevant to a given effect are required to provide support for your hypothesis, if both show a consistent response. **But note,** the chance of finding relevant studies where the response of interest was **not consistent** with your hypothesis will increase as you review more literature.

It is important that you justify how the study reported in the literature, is relevant to your investigation.

Note that the local data (which is generally evidence from a case study or study you have not yet published) and the literature evidence are treated the same way.

The transparency of the Causal Criteria framework requires that the literature reviewer documents the justification for the inclusion (or exclusion) of all studies deemed relevant. For example, justification may include a combination of geographical proximity, altitude, similarities in ecosystem processes and similar causal agent.

A study needs to have undertaken some statistical analysis, which provides a p-value, to be considered for inclusion. The exception is, where there is no p-value reported the effects must be obvious, e.g. mass extinctions.

The conceptual model will be relied upon heavily to guide the choice of 'relevant' studies (see below for guidelines on deciding relevance).

- What exactly is the agent of cause you are investigating?
- You may need to consider biogeography, ecology, etc. when deciding which studies are relevant.
- Are the temporal and spatial scales applicable to your investigation?

Quantifying the causal agent can be complex. For example, "flow change" may be a flow regime with many flow components that can be quantified. The flow regime components can range from coarse measures to more detailed flow components based on the known flow requirements of particular biota. Particular flow components can have very different consequences in different river systems, and on different biota. To use the Cotter River below Bendora Dam as an example, the flow components of interest could be defined as a reduction in the mean annual flow (MAF); reduction in frequency of all flood events < 1:10 year event and; reduction of baseflow. Use the following points to guide the reviewing process.

- Where **multiple causal agents** are identified, and reported separately in a body of literature, then their effect on different system components may be isolated. If it is not possible to isolate the different causal agents because the literature does not provide enough detail and the effect of non-relevant causal agents confound the effects of your particular agent of interest, then you must reject the evidence.
- **Record your search strategy** e.g. what key words or phrases did you use to search by and which databases did you search?
- **Specificity of association:** Did your literature search reveal a situation where a similar response was recorded in comparable studies but the stressor of interest was absent? Did your literature search and review reveal that a particular response occurs only with the causal agent of interest and not in relation to other

causal agents? Can you add further plausible causal agents to your conceptual model? For example, a similar change to the macroinvertebrate assemblage resulted from a seasonal change but the causal agent of interest to your investigation was not present. The specificity of the association needs to be documented.

- Species-specific evidence for species other than that under consideration should be rejected unless adequate information is provided in the literature (or reviewer has knowledge) to **ensure relevance to the question under investigation** – and the justification must be documented. For example, if the question relates specifically to reproduction success of Murray Cod, is it appropriate to use supporting evidence from a study on Golden Perch, considering the different types of eggs? If so, then justify the relevance.
- When reviewing the literature and **deciding relevance** for taxa other than species-specific evidence under consideration, we contend that we would have greater confidence in overseas evidence (or other geographically distanced evidence) if a biological effect is reported at a high taxonomic level (e.g. macroinvertebrate families rather than species). But where a different species response is used the **justification should be documented**.
- Remember - only the effect accepted as **relevant** will be used (and weighted) as evidence. **A study may include many effects from different lines of evidence** – some may be relevant and others not.
- The information required to **establish relevance** of the literature can generally be found in the abstract, results and methods sections of the papers, meaning that the entire paper may not need to be read.

Task check list:

- Search for relevant literature and record your search strategy
- Record details of literature reviewed (e.g. author, publication date, title, etc.). It is also important to retain a record of those studies deemed to be not relevant.
- Did the studies undertake statistical analysis providing p-values?
- Document the justification for including each study in your investigation. How is the study comparable or relevant to the investigation?
- Document the specificity of the association. Does the response occur only with the stressor of interest and not in relation to other non-relevant stressors? Can you add further plausible agents or stressors to your conceptual model?
- **Step 6** - After the literature review, **the conceptual model can be refined and the question clarified (i.e. re-visit Steps 3 and 4)**.

It may be helpful to record the following details for each effect in a spreadsheet as you review the literature.

Citation details	Causal agent	Effect	Study design type (see definitions in Step 7)	Number of independent sampling units - CONTROL or reference
Number of independent sampling units - IMPACT	Was the response consistent with your hypothesis/question (as stated in step 1)? Yes or No	Was there a dose response? Yes or No	Evidence of the causal agent or stressor found in the organism of interest? Yes, No or NA	Is the study relevant to your investigation? Yes or No
Justification for including (or excluding) the study	Does the response occur only with the causal agent of interest and not in relation to other relevant causal agents? Yes or No	Can you add further plausible causal agents to your conceptual model? If so, list them.		

Step 6

Refine conceptual model

Did the literature review identify additional potential causes and effects?

The conceptual model (Step 3) will need refining and also revisit Step 4 (Deciding on the relevant cause and effects) if the literature review has revealed additional plausible causes and effects.

Task:

- You will need to **review your conceptual model** (Step 3) and **re-visit Step 4** if the literature review has revealed **additional plausible causes and effects**
- Document these additions.

Step 7

Catalogue and weight the evidence

Evaluate the literature and the local data

Cataloguing and weighting process - overview

In the cataloguing process, for each effect the number and the importance of studies are recorded alongside the different causal criteria.

Note that the local data (which is generally evidence from a case study or study you have not yet published) and the literature evidence are treated the same way.

The Causal Criteria framework uses three main types of causal criteria to organize and weight the evidence (in terms of the **number of High and Low quality studies**) for the association between a given causal agent and effect (Table 1):

- Response
- Dose Response
- Consistency of Association

High levels of evidence for either or both the '**Response**' and the '**Dose response**' causal criteria show that an association between a causal agent and effect has been made (see definitions in Table 1). The association develops into a strong causal link if the expected response (e.g. a biological effect) is observed most of the time when the particular causal agent of interest occurs (i.e. the '**Consistency of Association**' criterion). Other causal criteria also provide support and strengthen confidence in your conclusion (Table 1).

The cataloguing and weighting process is explained in more detail in the section '<u>PROCESS FOR WEIGHTING EVIDENCE</u>'.

Table 1. Causal Criteria used in an investigation of causality (adapted from Downes et al. 2002). Note that the criteria are not necessarily listed in the order of their perceived importance.

Causal criterion - Levels of Evidence	**Description**	**Comments**
Plausibility	Is there a mechanism (e.g. biological mechanism) that could explain the relationship between the potential causal agent and the potential effect?	This component has been absorbed into our conceptual model rather than kept as a separate level of evidence, and is not used in the weighting process.
Presence of a **Response** (e.g. biological response)	There is evidence of the response following the potential causal agent (e.g. biological response)? This level includes results from all types experimental designs ranging from 'after impact only' to more complicated studies investigating natural or experimental gradients (see Table 3).	This is a recasting of the Downes et al. (2002) "experimental evidence" level, but includes evidence from all types of studies
Evidence of a **Dose Response** relationship with the causal agent	There is evidence of a dose response relationship between the agent of cause and the biological response, possibly from a study design using a natural or experimental gradient. A "Dose response" is also a "Response" but this criterion is a more compelling subset of the studies described for the 'Response' criterion above	As described by Downes et al. (2002) The relationship between the amount of exposure (dose) to some causal agent and the resulting response. A dose-response effect means that as the dose increases, so does the effect. The weighting for a "Dose Response" is the same as for "Presence of a response" until the study design is considered. Note that greater weight is given for studies with the power to establish such a response (e.g. Gradient response models and BACI designs).

Consistency of Association	A consistent spatial and temporal association of causal agent and effect. The expected response **ALWAYS** occurs in the presence of the causal agent.	As described by Downes et al. (2002). If the response or dose-response under investigation is not detected in a relevant study then the weight of that study will contribute toward a score for the lack of 'consistency of association' criterion. Thus, if a particular study is deemed relevant to the question under investigation (which implies that the causal agent is present) but no response was detected, then the association lacks consistency. Depending on the strength of the evidence it may provide support for a counter hypothesis.
Evidence of the causal agent found in biota	Evidence of the causal agent is **found in the organism** of interest, perhaps in the form of a chemical residue	Currently, this evidence is reported but not used in the weighting process. If evidence was found in biota it would provide support and strengthen confidence in your conclusion if that were the case.
Specificity of Association	Does the response (e.g. biotic response) occur **ONLY** with the causal agent of interest, and not in relation to other causal agents? It would provide support and strengthen confidence in your conclusion if that were the case. Documentation is required (yes, no, or NA). Specificity of Association is not treated as a separate causal criteria for weighting.	It is unlikely in ecological studies that a causal agent will be specific enough to answer positively to this question. It may be more useful in asking, ***can further plausible causal agents be added to the conceptual model***? Are there other plausible reasons for the response? If the literature review has revealed other plausible reasons then the conceptual model requires modification – the process should be documented and reported.
Agreement	Consistency of evidence	An output table should be

across effects	across the various effects	constructed to catalogue the evidence showing the number of lines of evidence (effects) where the overall conclusions are consistent with your hypothesis being tested, compared to total number of effects identified. This evidence is reported but currently not used in the weighting process.

Study type, control and impact locations

Important point: remember that the evidence you weight is the evidence that you are actually accepting as relevant. This could be only a part of the reported study, which may require a close look at the documented study design.

The study design type, number of independent sampling unit controls and number of independent sampling unit impact locations will all contribute to the overall study weight. The study design components of space and time are defined in a number of different ways in the literature so we defined some of these components below (Table 2), mostly according to Downes et al. (2002).

Table 2. Definition of some study-design terms used in the Causal Criteria framework

Term	Definition
Periods	Time is viewed as two major sampling periods, Before and After. Within these periods are time intervals, which can be of different temporal scales, e.g. monthly or yearly sampling.
Locations or independent sampling units	Locations are spatial units that may be in areas in which the same kind of human activity occurs (i.e. true replicates of the impact), or independent areas that serve as comparisons to impacted areas (i.e. replicates of un-impacted areas). Locations are presumed to be spatially distinct to the extent that conditions at one location are independent of conditions at another location.
Replicate	A replicate is the same treatment in a different location. We consider replication at the location level. For instance, replicate locations can be in different rivers or catchments. There can also be replication in time if the sampling time intervals (see period above) are large enough. Samples through time from one location can become replicates instead of subsamples because the large intervals mean that the samples will tend to represent independent and random

	observations (part of the formal analysis in such cases may include preliminary tests for temporal autocorrelation).
Subsamples	Subsamples are observations within a location. The first level of subsamples is sites within the impact location. There can be further subsamples within the sites depending on the study. Subsamples can also be observations from the same location on multiple occasions with small sampling time intervals (i.e. can be assumed or shown to be temporally autocorrelated).
Sites	Sites are the first level of subsamples from a (control/reference or impact) location.
Reference location	Reference locations are areas that are as close as possible to the state of the environment undisturbed by human activity. Reference locations are not chosen with a particular impact in mind but to represent what other locations could be like in the absence of human disturbance.
Control location	Control locations are areas that are as similar as possible in all respects to the impact location, except for the presence of the agent or stressor. The intention is to use the control locations to isolate the effect of the particular human activity from a range of other processes. Under some circumstances, when a human activity occurs in an otherwise undisturbed area, control and reference locations may have the same attributes. However, when a new activity is contemplated for an area that has already been highly modified, the controls should be locations that are themselves modified in similar ways.
Counting independent sampling units (i.e. impact and control locations)	In the Causal Criteria framework when we count independent sampling units i.e. impact locations and control locations. We do not count subsamples because they do not increase inferential power in the way that replicate locations do. This is illustrated graphically below. In the illustration below there are two sites on the main stem of a river below the dam and three subsamples at one of the sites. These subsamples represent a single independent sampling unit or impact location because the three subsamples are not independent. Similarly, there are only two control locations even though two sites were sampled on one of the tributary streams.

In addition, it is important to consider the number of controls or impact locations actually used in the analysis of each effect because in some cases not all of the control locations reported in a paper are used for the analysis of every effect reported. It may help to check the degrees of freedom for each result you actually accept for a relevant effect, (**the result that you actually weight when using the Causal Criteria method)**, it is

that particular result you need to consider when counting the number of impact and control/reference independent sampling units or locations for each effect.

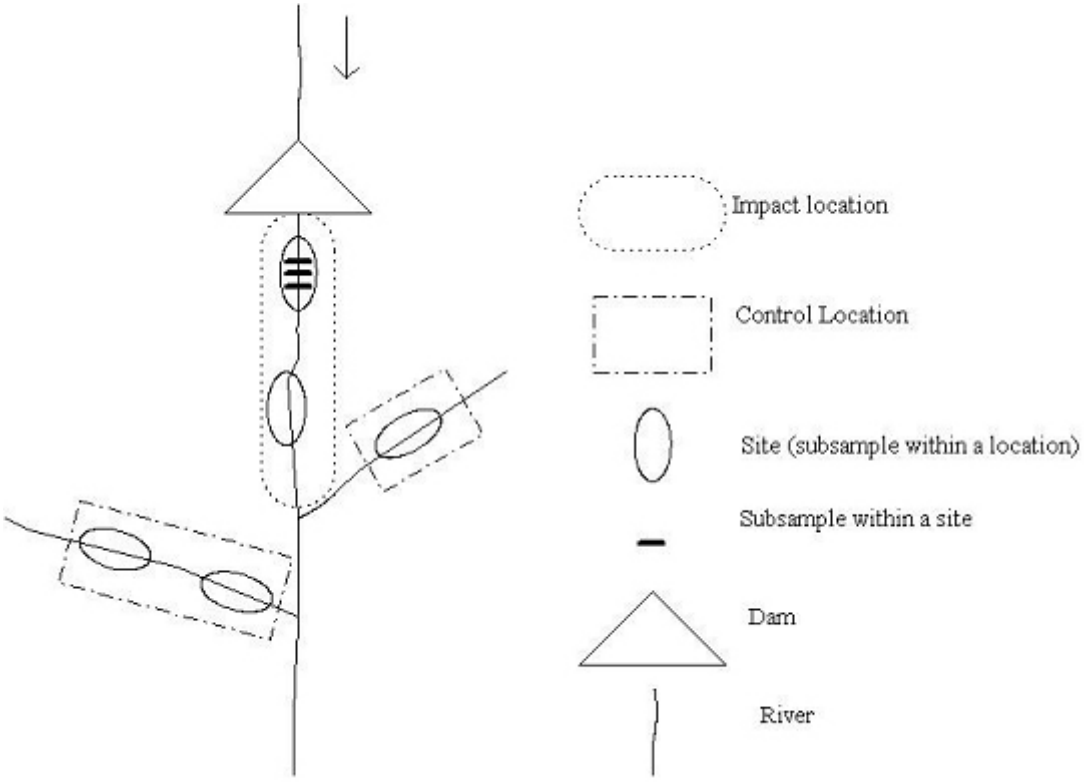


Figure 3. Some study design components of space (locations, sites and subsamples) in a scenario where the independent sampling unit is a location below a dam (the one impact location has two sites and subsamples within the sites) is being compared to two similar independent sampling units unaffected by a dam (i.e. 2 control locations).

Study design types

The studies are weighted differently depending on their inferential power, thus, different study design types contribute more or less to the overall study weight (Table 3).

Table 3. Definitions for the study types used in the Causal Criteria framework.

Study design type	Default weight	Description
After impact only	1	A comparison of variable(s) from impact locations with a standard of some kind but no control/reference location sampled and no before data compared e.g. DO measured at impact location and compared to a standard.
Control/Reference vs. Impact - no before	2	Any case where reference/control locations are being compared to impact locations without before data for the impact locations. This may include cases where locations are compared to output from models based on reference/control locations (e.g. RIVPACS or AUSRIVAS). In such a case the number of impact locations will be the number of impact locations sampled and compared to the model whereas the number of control locations will be set to any value that will give maximum weight for the number of controls study quality attribute. The numbers of control locations are counted in this way to acknowledge the enhanced inferential power from the number of reference/control locations in the predictive model. The reviewer needs to check or assume that the predictive model is robust and has been cross-validated.
Before vs. After - no reference/control	2	Sampled Before and After onset of agent / stressor with no reference/control.
Gradient Response Models	3	An investigation of an association between the agent / stressor and response along a gradient of the agent or stressor, e.g. using correlation or regressions. The data may include pristine locations through to highly disturbed locations or may have no pristine locations but have locations disturbed to different degrees. The number of control locations and the number of impact locations are set to a value that will give maximum weight to the control and impact study quality attributes to take into account the greater inferential power from a dose response gradient.
BACI/BARI	4	Before After Control Impact or Before After Reference Impact. 1 control/reference location, 1 impact location sampled once (may include subsampling within a location). Compares changes at

		two locations, i.e. Control/Reference and Impact locations, Before and After the impact.
BACIP	4	1 control location, 1 impact location sampled through time at the same times, i.e. paired measurements from control and impact locations. Sampling through time with large intervals during both the Before and After periods is used to estimate the temporal variation in the differences between control and impact locations. The temporal variation is used to assess the average difference Before and After the activity commences. The time intervals are large enough to prevent autocorrelation in these differences, and are viewed as a random sample of possible values in each time period.
MBACI/MBARI	4	Either Multiple Control/Reference locations and 1 or many Impact locations. Either multiple impact locations and 1 or many control/reference locations. Also includes replication in time.

For any study to be considered for inclusion using the Causal Criteria protocol, some statistical analysis providing a p-value is required. The exception to this rule is where the effects are obvious e.g. mass extinctions.

Task

At this stage in the Causal Criteria application reviewers are required to document the study details (e.g. study type, number of impact and control independent sampling units or locations – see definitions above) that are used to weight the evidence and to determine the quality of the evidence documented in the literature or local data – remember that the local data are treated the same way as the literature evidence. Note that the Causal Criteria Analysis Software program has been designed to make this task easier.

- **Specify the study type** relating to each relevant cause and effect. The different study types are assigned different weights. For example, if a study uses a Before-After-Control-Impact (BACI) design, the maximum study weight value of 4 will be assign if using the default values. A study may include evidence for many effects, some may be relevant to your question and others not, this situation may require a closer look at the study design detailed in the literature. A study type should be assigned to each cause and effect identified from the literature (or local case study) not the entire study if not all of the evidence is relevant. **Only apply the study type to the result you will use (and weight) as evidence**.
- **Specify the number of independent sampling units** (i.e. reference or control locations). Having a control brings an improvement in inferential power (Downes et al. 2002). Again, if using the default values, the maximum default 'weight' value of 3 will be assigned if 2 or more reference/control locations are used for a given cause and effect.

- **Specify the number of independent sampling units for impacts or impact locations**. A larger number of impact locations leads to a better estimate of the range of dynamics that might be experienced by the impact locations and the control conditions with which they are being compared. The maximum default 'weight' value is 3 if 2 or more impact locations are used for a given cause and effect.
- **Overall study weight** - For each piece of cause and effect evidence identified from a study, the weighting value for each of the above three categories are summed to give an overall study weight. If you are using the CCA Software the program will do this for you.
- The scores can also be converted to quality categories. The advantage of using study quality categories is that it more immediately relays the importance of the study than a numerical value. For example, the default values for overall study weight are 5-10 = **High quality** and 1-4 = **Low quality** study. Again the available CCA software can make this job easier.

Process for weighting evidence

Overall study weight

Studies are weighted on three characteristics to provide an overall study weight:

1. **type of study design;**
2. **number of independent sampling units used as controls; and**
3. **number of independent sampling units used to investigate impacts.**

Type of study design

Justification of the weighting: Studies in which error terms are well controlled (e.g. BACI designs) should exert greater influence than less rigorously controlled designs (e.g. only impact locations sampled). Therefore, these studies have more weight than studies for which we lack information, such as after impact only (Table 4).

Table 4. Study design types and default weight values

Study design type	Weight
After impact only	1
Reference/Control vs. impact (no before)	2
Before vs. after (no reference/control)	2
Gradient response model	3
BACI or BARI MBACI or Beyond MBACI	4

Number of control independent sampling units

Justification: Having a control brings an improvement in inferential power (Downes et al., 2002). There is some increase in inferential power derived from having more than one control. A larger number of control locations or sampling units is important because it better estimates the envelope of 'normal' behavior (Downes et al., 2002) so that departure from 'normal' can be detected with more confidence. Therefore, studies with 2 or more controls have more weight than studies with no controls (Table 5). Studies that use a predictive model for the reference condition (such as RIVPACS and AUSRIVAS that are based on many reference sites) should be entered as having at least 2 control locations and given the highest weight.

Table 5. Number of control locations & proposed weights

Number of controls / reference locations	Weight
0	0
1	2
>2	3

Number of independent sampling units for impacts

Justification: A larger number of impact locations or test sites leads to a better estimate of the range of dynamics that might be experienced by the impact locations and the control conditions with which they are being compared. Therefore, studies with more than 2 impact locations have more weight than studies with only 1 impact location (Table 6).

Table 6. Number of impact locations & proposed weights

Number of impact locations	Weight
1	0
2	2
>2	3

Overall study weight - calculated from the above 3 tables

The weight values for the above three categories for each study are summed to give an overall study weight. We considered that at this point it was useful to convert these scores to quality categories. The advantage of using study quality categories is that it more immediately relays the importance of the study than a numerical value.

A study that meets the minimum criteria for a satisfactory study would at least have an 'after-only' study type, no control locations and 1 after impact location, resulting in a minimum study weight of 1 (i.e. 1 + 0 + 0).

A much better study would have a reference versus impact study type, 1 control location and 1 impact location. The weight for that study would be 4 (i.e. 2 + 2 + 0). That study would still be considered a low quality study because it has no replication.

A study weight above 4 would mean that the study at least has a reference versus impact study type and at least control location replication or impact location replication. Such studies would have a maximum weight of 10 and have been classed as high quality studies because their inferential power is greatly increased by the presence of replication (Table 7).

Table 7. Scheme for categorizing the importance of each study

Overall study weight	Study importance category
5-10	High quality
1-4	Low quality

Combining studies and causal criteria

In the Causal Criteria framework three types of causal criteria are used to weight the relevant literature evidence (in terms of the number of High and Low quality studies) for a cause and effect association.

1. **Response**
2. **Dose Response**
3. **Consistency of Association**

High levels of evidence for either or both the '**Response**' and the '**Dose response'** causal criteria show that an association between the causal agent and the effect has been made in relevant works. The association develops into a stronger causal link if the expected effect is observed most of the time when the causal agent occurs (i.e. the '**Consistency of Association**' criterion).

For the '**Response**' and '**Dose response**' causal criteria, **combinations** of High and Low quality studies with a summed quantitative study weight of **20 or more** are deemed to have a **HIGH level of support** (from the literature and accompanying studies). A cause and effect with a summed quantitative study weight of **less than 20** has a **LOW level of evidence** for the '**Response**' and '**Dose response**' causal criteria (Table 8).

Note: In Table 8 below, the values representing the number of low and high quality studies were selected at random and are used as an **example only**. Also, a median study weight is used here for the purpose of the example (2.5 for low quality studies and 7.5 for high quality studies – from Table 7), however, the actual overall study weights (based

on study design, control and impact locations) for each of the studies would be used in a real situation.

Table 8. Example of decision rules for **Response** and **Dose response** causal criteria.

Number of low quality studies showing support	**and**	**Number of high quality studies showing support**	**Sum of the quantitative study weights (2.5 = Median low quality study weight; and 7.5 = Median high quality study weight – see Table 8)**	**Conclusion – evidence for causal criterion**
0	and	3	(0 x 2.5) + (3 x 7.5) = 22.5	High
2	and	2	(2 x 2.5) + (2 x 7.5) = 20	High
1	and	2	(1 x 2.5) + (2 x 7.5) = 17.5	Low
7	and	0	(7 x 2.5) + (0 x 7.5) = 17.5	Low

For the '**Consistency of Association**' criterion, combinations of High and Low quality studies with a study weight of **20 or more** shows a **LACK of consistency** for the cause and effect relationship (Low level of support), whereas a study weight of **less than 20** shows **HIGH Consistency of Association** for the cause and effect relationship (High level of support) (Table 9).

Table 9. Example of decision rule for the '**Consistency of Association**' causal criteria. Note that a comparable study with the relevant causal agent present but a response not consistent with your hypothesis will contribute to a lack of consistency. Study weight of **20 or more = Low** level of support), whereas a weight of **less than 20 = HIGH** level of support.

Number of low quality studies showing lack of consistency	**and**	**Number of high quality studies showing lack of consistency**	**Sum of the quantitative study weights (2.5 = Median low quality study weight; and 7.5 = Median high quality study weight)**	**Conclusion – evidence for causal criterion**
0	and	3	(0 x 2.5) + (3 x 7.5) = 22.5	Low (Lack of consistency)
2	and	2	(2 x 2.5) + (2 x 7.5) = 20	Low
1	and	2	(1 x 2.5) + (2 x 7.5) = 17.5	High
7	and	0	(7 x 2.5) + (0 x 7.5) = 17.5	High

The possible outcomes for a given causal relationship are show below (Table 10). The conclusions show whether the evidence, according to the causal criteria, leads to enough support for causal relationship between a given causal agent and effect, or whether there is no support, insufficient evidence or inconsistent evidence for a causal relationship. The minimum requirement for demonstration of a causal relationship is for the support for both "Response" and "Consistency" to be HIGH (i.e. outcomes 1 and 2, in Table 10).

Table 10. Possible outcomes depending on the strength of evidence for the causal criteria of a given cause and effect relationship. H = High; L = Low.

Outcomes	Response	Dose Response	Consistency of Association	Conclusion
Outcome 1	H	H	H	Support for hypothesis
Outcome 2	H	L	H	Support for hypothesis
Outcome 3	L	L	H	Insufficient evidence
Outcome 4	H	H	L	Inconsistent evidence
Outcome 5	H	L	L	Inconsistent evidence
Outcome 6	L	L	L	Support for counter hypothesis

People may wish to set their own weights (note that the software package also allows for that). However, we have provided our justification for setting the default values as they are, and if users wish to change them they can (the weights used are reported in the final output report provided by the software) but users should also provide a justification for the change.

To incorporated flexibility, we have also retained ability for people to include other causal criteria. This flexibility is not currently in the weighted, quantitative assessment of evidence, but in the final 'verdict'. The final conclusion is always up to the user even when using the software package. However, ignoring the evidence provided to accept a different verdict would require justification to be documented because we see a major strength of this Causal Criteria approach is in its transparency.

When assembling the catalogue, the resolution of the 'effects' will need to be specified. For example, will "a decrease in fish taxa richness" and " reduction in fish abundance" be combined to form "Change to fish assemblage" evidence? The appropriate resolution will be study specific – let the conceptual model guide your decision.

Step 8

What is the verdict?

Accept or reject your hypothesis?

Data from the literature review, and local data if used, can now be assembled into an evidence catalogue, which can then used to accept or reject your hypothesis, or to assess if you have enough evidence to accept or reject the hypothesis.

- The evidence catalogue should show the number of High quality and Low quality studies supporting the three main causal criteria (i.e. 'Response', 'Dose Response' and 'Consistency of Association').
- Convert these numbers to quality categories – see the section on '**Process for Weighting Evidence**' for the default decision rules for assigning High and Low evidence support for the three causal criteria or let the software program do this for you.
- The evidence catalogue will also show support from the other causal criteria that are not used in the weighting process but could also strengthen confidence in the verdict.

The evidence for each different effect and the three main causal criteria (Response, Dose Response, Consistency) should then be aggregated to arrive at a conclusion about the total level of support for your hypothesis (Table 10).

- The various outcomes show whether the evidence according to the causal criteria leads to enough support for a particular causal relationship or whether there is either no support or insufficient evidence for the causal relationship.
- The minimum requirement for demonstration of a causal relationship is "Response" and "Consistency" to be HIGH for at least one of the causal relationships.
- At this stage, a 'magic formula' has not been created to mechanically proceed from Table 10 to a conclusion about your hypothesis. The questions to be addressed are too variable to permit this level of automation.
- The final report will provide transparency to show the logic and evidence used to reach your verdict.
- Some questions may require evidence across many causal relationships in order to infer causality.
- **Agreement across effects**: Construct a table similar to Table 10 for each effect (the software can do this for you) to show if there was agreement across the different effects. If so, this can provide yet further support your conclusion.

Reporting conclusions

There are four different conclusions that can be arrived at for final reporting on the evidence collected:

1. **Support for hypothesis;**
2. **Insufficient evidence;**

3. **Inconsistent evidence; or**
4. **Support for counter hypothesis.**

Support for hypothesis

A conclusion of 'support for hypothesis' provides a strong scientific basis for making management decisions. Based on the Causal Criteria assessment, managers can conclude with a degree of confidence that the given human activity or natural event in question will have a particular affect.

Insufficient evidence

A conclusion of 'insufficient evidence' may indicate a knowledge gap with respect to the potential causal relationship in question.

The first step would be to conduct a more extensive literature search, to find more articles that address the question being asked. If further information is not available even after an extensive literature search then the knowledge gap could be addressed by conducting studies designed to gain a better understanding of the problem in question. Such studies would provide valuable information for the future environmental management so be sure that the new study scores highly in the overall study weight. If you are managing in the presence of knowledge gaps an adaptive management approach is recommended. The study weighting components should be used as a guide to study design.

Inconsistent evidence

A conclusion of 'inconsistent evidence' means that although there are studies that show a response to the causal agent, which is consistent with your hypothesis, there are also several studies that do not show the expected response even though the potential casual agent was present.

The first step is to review the relevance of the literature used in the Causal Criteria analysis:

Have you fully documented the justification for including all the studies used and their relevance to the question asked? Anyone evaluating your conclusions would look there first to assess your understanding of the situation.

1. **Have you weighted the studies correctly (regarding study type, number of impact and control locations)?**
2. **Have you aggregated the evidence appropriately (i.e. grouped effects appropriately)?**
3. **Re-visit your conceptual model. Have you included all plausible causal relationships?**
4. **Is there some aspect of those studies that show a response, which makes them different from the studies that do not detect a response?**
5. **What other confounding factors may have been present?**
6. **What taxonomic resolution was used? Was it appropriate?**

If no changes to the Causal Criteria assessment are made, a knowledge gap may exist (see comments under 'insufficient evidence' heading above).

Support for counter hypothesis

A conclusion of 'support for counter hypothesis' means the supporting evidence for the causal relationship is low; and a **number of studies show a lack of support for 'consistency of the association'**. A number of studies showed that the **response** or **dose-response** under investigation was not detected in the relevant studies (even though the causal agent was present), so the weight of those studies contributes toward a score for lack of '**consistency of association**'. The combination of LOW scores for all three of the major causal criteria used to weight the evidence (response, dose-response and consistency of association) indicates support for the counter hypothesis i.e. the Causal Criteria analysis concludes that the causal agent does not always produce a response as hypothesized.

No evidence supplied

The words 'no evidence supplied' should be reported when no evidence is provided for a given potential cause and effect or associated individual causal criteria (i.e. no relevant studies reviewed to provide evidence for the causal criteria).

Task:

- Assemble the catalog of evidence (note, the Causal Criteria software program can output the various tables of evidence).
- Report on each Step of the framework (the software can also provide this report).
- Record the verdict. Accept or reject the hypothesis based on the evidence you have collected.

Causal Criteria EXAMPLE

In this fictitious example

A river manager needs to determine whether reducing the mean annual flow downstream of a dam will have adverse ecological impacts

- For this example, the literature review identified five papers (note that you would generally review more than 5), which have the same causal agent and the studies were undertaken in close geographical proximity to the dam in question. The 'effect' in each of the studies involved macroinvertebrates as the biological indicators of ecological health.

- Details of the study design, number of control sites and number of impact sites in each of the five papers are provided below (Table 12). The overall study weight (Table 13) has been calculated using the method described in Chapter 3 'PROCESS FOR WEIGHTING PAPERS'.
- The causal agent is a 'reduction in mean annual flow'.
- The effect under consideration for this example (although there may be more) is 'a decrease in macroinvertebrate taxonomic richness'.

The Hypothesis is that **a reduction in mean annual flow downstream of dams will result in reduced macroinvertebrate taxa richness**.

In this example, all of the studies are comparable (i.e. **relevant to the investigation**) because they have the same causal agent and they are in close geographical proximity. The first four papers found a response (i.e. **shows support for the hypothesis**), however, paper 5 did not find the expected biological or dose response in the macroinvertebrate assemblages to reduction in mean annual flow (Table 14).

The study weights for each of the papers showing support for the hypothesis are summed (i.e. Papers 1 – 4). Study weights of less than or equal to 19 are deemed to have a high level of support from the literature (Table 15). None of the papers looked at Dose-response relationships.

The sum of the study weights for papers **NOT showing a response** (i.e. paper 5) is calculated for the '**consistency of association**' criterion (Table 16).

Overall, there was support for the hypothesis that reduction in mean annual flow causes a decrease in macroinvertebrate taxonomic richness (Table 17).

Table 12. Details of the study design, number of control sites and number of impact sites for five fictitious papers.

Literature	Study design	Number of control sites	Number of impact sites
Paper 1	BACI	3	3
Paper 2	Reference/Control vs. impact (no before)	2	4
Paper 3	After impact only	0	2
Paper 4	Reference/Control vs. impact (no before)	1	6
Paper 5	After impact only	0	5

Table 13. The overall study weights for five fictitious papers.

Literature	Study design weight	Control sites weight	Impact sites weight	Overall study weight	Study importance category 1 - 4 = Low importance; 5 - 10 = High importance
Paper 1	4	3	3	4 + 3 + 3 = 10	High
Paper 2	2	3	3	2 + 3 + 3 = 8	High
Paper 3	1	0	2	1 + 0 + 2 = 3	Low
Paper 4	2	2	3	2 + 2 + 3 = 7	High
Paper 5	1	0	3	1 + 0 + 3 = 4	Low

Table 14. The biological response and relevance of five fictitious papers.

Literature	Response	Dose response	Comparable (or relevant) Note that justification for inclusion of each study should also be documented and reported
Paper 1	Yes	Na	Yes
Paper 2	Yes	Na	Yes
Paper 3	Yes	Na	Yes
Paper 4	Yes	Na	Yes
Paper 5	No	Na	Yes

Table 15. Response weighting for the five fictitious papers.

Number of low quality studies showing support		Number of high quality studies showing support	Sum of the quantitative study weights(Study weights of 20 or more are deemed to have a HIGH level of support from the literature)	Conclusion – evidence for causal criterion
1	AND	3	(1 x 3) + (1 x 10) + (1 x 8) + (1 x 7) = 28	High

Table 16. Decision rule for the 'Consistency of Association' causal criteria

Number of low quality studies showing lack of consistency		Number of high quality studies showing lack of consistency	Sum of the quantitative study weights (less than 20 shows HIGH consistency of association in the line of evidence)	Conclusion – evidence for causal criterion
1	AND	0	(1 x 4) + 0 = 4	High

Table 17. Causal Criteria Analysis outcome for the causal relationship.

Biological response	Dose response	Consistency	Conclusion
High	No evidence supplied	High	Support for hypothesis

Note that for simplicity this example has used evidence for only one effect (**reduced macroinvertebrate taxa richness**). If more were to be used you would report each one as in Table 17 and assess for **'agreement across different effects'** (yet another causal criterion to use as evidence).

GLOSSARY

Term	Definition
Adaptive management	Evaluating the performance of new management approaches and changing practices over time as experience is gained.
Causal Agent	A substance or activity that exerts some force or effect.
Bioassessment	Evaluation of the condition of an ecosystem that uses biological surveys and other direct measurements of the resident biota.
Response	There is evidence of the response following the causal agent. This includes results from all types of experimental designs from after-impact-only to more complicated studies investigating natural or experimental gradients.
Conceptual model	A depiction or representation of the most current understanding of the major ecosystem features and processes (including biological, physical, chemical and geomorphic components) of a particular environment.
Consistency of association	Consistent spatial and temporal association of causal agent and response. The expected response ALWAYS occurs in the presence of the causal agent.
Dose response relationship	The relationship between the amount of exposure (dose) to some causal agent and the resulting response. A dose-response effect means that as the dose increases, so does the effect.
Hypothesis	A theory or assumption that can be tested by further investigation.
Impact location	The impact location is a term used to describe a site potentially impacted by a given causal agent but is in unknown condition (also known as test site).
Potential Effect	An system attribute that is investigated in relation to a causal agent e.g. tadpole abundance, macroinvertebrate species richness, macrophyte biomass, number of fish abnormalities
MLLE	Multiple lines and levels of evidence.
Stressor	Any physical, chemical or biological entity that can induce an

	adverse response.
Study design	The science and art of planning how an experiment will be conducted to get the most valid and reliable results (also referred to as research or experimental design). There are several types of study designs.
Treatment	Controlled technique or action applied in a specified process or experiment.

Acknowledgements

This document builds on prior work by the Cooperative Research Centre for Freshwater Ecology (CRCFE).

References

Downes, B.J., Barmuta, L.A., Fairweather, P.G., Faith, D.P., Keough, M.J., Lake, P.S., Mapstone, B.D. and Quinn, G.P. (2002). Monitoring Ecological Impacts: Concepts and Practice in Flowing Waters. (Cambridge University Press.) 446 pp.

Norris, R., Liston, P., Mugodo, J., Nichols, S. Quinn, G., Cottingham, P., Metzeling, L., Perriss, S., Robinson, D., Tiller, D. and Wilson, G. (2005). Multiple Lines and Levels of Evidence for detecting ecological responses to management intervention. In I.D. Rutherfurd, I. Wiszniewski, M.J. Askey-Doran and R. Glazik (Eds), Proceedings of the 4th Australian Stream Management Conference: linking rivers to landscapes, (pp. 456-463). Department of Primary Industries, Water and Environment, Hobart, Tasmania.