

The Collaboration for Environmental Evidence

A checklist for Editors and Peer Reviewers of Evidence Reviews

What is this checklist for?

Since 2018 the Collaboration for Environmental Evidence (CEE) has been collating a database of evidence reviews, published globally, of relevance to environmental management as a free service to organisations with evidence needs who want their decision making to be evidence informed. Hundreds of such reviews are published annually in a broad range of journals and organisations. As part of our evidence service CEE critically appraises each review for its reliability (risk of bias) in terms of conduct and transparency of reporting as well as limitations of the primary data. Our initial survey of reviews suggests that the reliability of reviews is, on average, very low and we would like to help to improve this situation and so better inform environmental decisions. The good news is that many problems with the conduct and reporting of evidence reviews are easily addressed and this checklist is intended to help authors, editors, and peer reviewers make or suggest basic improvements.

What does this checklist do?

This checklist provides a point by point appraisal of the confidence that can be placed in the findings of an evidence review by assessing the rigour of the methods used, the transparency with which those methods are reported and the limitations imposed on synthesis by the quantity and quality of available primary data. The purpose of the checklist is to assist editors and peer reviewers in improving the reliability of evidence reviews intended to inform decision making in the environmental sector.

How should this checklist be used?

This checklist can be used for any review or synthesis in the environmental sector that claims to provide a quantitative measure of the effectiveness of an intervention or the impact of exposure to human activities. This checklist may also be used more broadly for reviews of cause and effect questions. It is offered as an additional tool and is not intended to replace any established procedures or checks a journal may have in place.

‘Yes’ to all checklist questions (some may be not applicable) are expected for a reliable review or synthesis (see guidance notes below). Where ‘No’ is selected, editors and peer reviewers may wish to ask authors for revision.

Checklist Questions (see explanatory notes below for guidance)	Yes/No/NA
1. Review Question	
Is the review question or hypothesis clearly stated, and are key-elements defined (e.g. population, intervention/exposure, comparator, outcome)?	

Is the scope (e.g. geographical, taxonomic, temporal) of the question clear?	
2. General methods	
Does the review include a defined methods section covering all the synthesis stages conducted and providing sufficient detail to enable the method to be replicated?	
3. Search Strategy	
Are all search terms and search strings, with Boolean operators ('AND', 'OR' etc.) and wildcards, clearly stated for each major source (e.g. databases, search engines) so that the exact search is replicable by a third party (search terms for minor sources (e.g. specialist websites), if used, may be simplified)?	
Is there information about the dates of search and any other limitations justified (e.g. languages, no grey literature searches)?	
Are the sources searched comprehensive and capture both conventionally published scientific literature and grey literature using a combination of databases, search engines and specialist websites (may also be informed by stakeholders) or are limitations are fully justified?	
4. Screening articles and including studies	
Are study eligibility criteria precisely defined so as to be replicable (e.g. reliance on broad and potentially ambiguous terms should be avoided) and expressly related to each key element of the question (other criteria such as study design may also be considered)?	
Are eligibility criteria independently applied by more than one reviewer to a sample of justified size of the screened articles/studies at title, abstract and full text?	
Is consistency of eligibility decisions measured and reported and all disagreements between reviewers discussed so that the resolutions informed subsequent assessments?	
Is the number of unique articles found during the searches reported with the number of articles excluded at each stage of the screening process (maybe as a flow diagram)?	
Is a full list of eligible (included) articles/studies provided (not just included in reference list)?	
5. Critical appraisal of study validity	
Is an effort is made to identify all sources of bias relevant to each of the individual included studies (threats to internal and external validity)?	
6. Data Extraction	
Do the authors state the type of data to be extracted and the methods by which data from each study were extracted so that the process can be replicated?	
Are all extracted data reported in a table or spreadsheet so that the synthesis can be replicated? This includes the data used in the synthesis from each primary study (e.g 'raw' outcome metrics: means and variance measures) and meta-data.	
Is an explanation provided of how data extraction was cross checked for consistency between two or more reviewers?	
7. Data synthesis	
Is the choice of synthesis method (i.e. quantitative meta-analysis or narrative synthesis) is described in sufficient detail to be replicable and justified on the	

basis of characteristics of included studies, taking into consideration variability between studies in sample size, study design, context, outcomes etc?	
If meta-analysis is conducted, are statistical estimates of findings presented using a method that justifies approach (e.g. using study weighting and subgroup analysis)?	
If meta-analysis is conducted, is consideration given to study independence (e.g. through sensitivity analysis) and bias (e.g. tests for publication bias)?	
If meta-analysis is conducted, are potential effect modifiers (variables other than the factor of interest e.g. taxa being considered, location, habitat type, study design etc.) investigated statistically through meta-analysis (alternatively, evidence for heterogeneity in effects between studies is tested and reported as non significant)?	
8. Review limitations	
Is an explicit section or identifiable passage of text devoted to the authors' consideration of limitations of the conduct of the review process as well of the primary research/data?	

Explanatory notes

1. THE REVIEW QUESTION

A well-defined question (or hypothesis) is crucial for assessing the reliability of subsequent decisions on searching and screening for eligible studies, as well as forming the basis for critical appraisal of study conduct and for data extraction and synthesis. Cause and effect questions usually require definition of the population (biological or statistical) of interest, the causal factor (exposure or intervention), the comparator (e.g. no or alternative intervention and the outcome (measure of effect)

2. THE METHOD

As with all scientific studies, the methods should be sufficiently reported so as to be replicable. This is equally true for reviews that aim to synthesis primary data.

3. SEARCHING FOR STUDIES

An optimal search for literature should possess three key properties: comprehensive (maximises the number of potentially relevant studies found), systematic (avoiding ad hoc search strategies reduces the susceptibility to bias resulting from e.g. no defined endpoint of search or 'cherry picking') and transparent (readers should be able to replicate and evaluate the search).

Is the approach to searching clearly defined, systematic and transparent?

Search strategies should be outlined in the predefined protocol or review methods. An optimal search for literature should aim to maximise comprehensiveness (aiming to identify all relevant studies) and transparency (readers should be able to replicate and evaluate the search). This is to avoid 'cherry-picking' studies or assembling a biased or unrepresentative body of evidence. Where possible, advice should be sought from an expert such as an information specialist/scientist.

Is the search comprehensive?

The resources used to find relevant literature influence the comprehensiveness and reliability of the synthesis. The principal sources for locating peer-reviewed articles are electronic databases of scientific literature and academic search engines, with a range of supplementary methods. No single database indexes all peer-reviewed articles. Moreover, these sources are unlikely to capture potentially relevant grey literature (e.g., reports by governmental and non-governmental organisations, unpublished studies) and consequently can be complemented by searching thesis repositories, websites of relevant organisations and conducting internet searches. Other supplementary search strategies include citation chasing (backwards and forwards), and contact with experts in the field.

4. INCLUDING STUDIES

Comprehensive searches may generate a large number of articles that vary widely in their relevance to the synthesis. Authors must then determine whether or not each article is sufficiently relevant (eligible) for inclusion in the data synthesis stage. However, the choice of eligibility criteria can influence the conclusions of the synthesis, and the application of inadequately defined criteria can be subjective and lead to biases. Decisions over which studies are relevant for inclusion should therefore be based on clearly defined criteria, and should be replicable and transparent. Criteria 4.1-4.3 refer only to studies included/excluded on the basis of relevance to the review question – see 5 for inclusion/exclusion on the basis of methodological quality.

Are eligibility criteria clearly defined?

Clearly stated criteria for eligibility decisions minimise the potential for subjective decisions to influence which studies are included in the review, increase the transparency of the synthesis, and allow readers to assess the validity of the criteria to the review question. In addition to following the review question, eligibility criteria may define limits on the type of primary research to be considered in terms of (for example): geographic scope, type of data reported, type of intervention or impact, study design, date.

Are eligibility criteria consistently applied to all potentially relevant articles and studies found during the search?

More than one person should screen studies for inclusion to reduce the risk of human error and to ensure that the criteria are applied consistently to the articles returned by the search. If more than one person independently evaluates the relevance of the same articles, the consistency of

inclusion/exclusion decisions can be assessed. Piloting the criteria, and discussing and refining the eligibility decisions can also ensure they are consistently applied.

Are eligibility decisions transparently reported?

Listing all articles that were screened for eligibility and indicating whether each was included or excluded in data synthesis (usually as supplementary material), makes it clear whether potentially relevant studies have been omitted according to the eligibility criteria or were not captured by the search. Documenting the reasons for article exclusion at full text is essential for transparency and replicability.

5. CRITICAL APPRAISAL

Primary research can vary widely in methodological validity (internal validity) and study context (external validity). Internal validity can influence the findings of the research, and, if not properly accounted for, the conclusions of syntheses that use it. External validity can influence the relevance/applicability of the study to users of the findings in individual contexts. Critical appraisal involves transparently evaluating the design and conduct of each included study based on methodologies, which can then help to objectively account for variation in study quality by placing greater emphasis on the most reliable studies.

Does the review critically appraise each study?

Documented critical appraisal, as applied to each individual each included study, using relevant, pre-defined critical appraisal criteria allows the author(s) of the synthesis and the reader to make more objective assessments of the relative reliability (or weighting) of each study. Some potentially relevant studies may not meet baseline methodological requirements (e.g. small sample size, pseudoreplication, spatial autocorrelation, lack of appropriate controls etc.) and so may be excluded from the synthesis. Effectively, these studies are weighted as 'zero'. Studies included in the synthesis may be treated differently according to the rigour of the sampling design, according to differences in sampling effectiveness (e.g. sample size, sampling area, study duration, etc.), or according to their generalisability for the synthesis in hand (e.g. spatial scale, study setting, etc.).

6. DATA EXTRACTION

The volume and type of data collected by primary research articles varies substantially, even when similar questions are addressed. Authors of evidence syntheses must make decisions on which data to extract and on how to extract this information. These decisions may influence the findings of the synthesis, and so to minimise bias, the approach to data extraction should be clearly stated and, wherever possible, the extracted information should be comparable and consistent between studies.

Is the method of data extraction fully documented?

Transparently identifying a consistent set of data to extract from each study, for example into a structured data extraction sheet, allows the process to be replicated and evaluated by a third party, and reduces the potential for bias over which data are extracted from individual studies. Typically, extracted information from each study included in the review comprises: study aims; intervention details, study design; population characteristics; comparator details and results (point estimates and measures of variance).

Are the extracted data reported for each study?

Providing a summary in which the population, intervention/exposure and outcome for each study are stated makes data extraction transparent, and makes it easier for readers to locate the most relevant primary literature and conduct supplementary analyses if required. Data may be provided in additional files or in an open access repository.

Were extracted data cross checked by more than one reviewer?

Checking data extraction improves accuracy by ensuring the correct data are extracted for each element and reduces the risk of errors due to interpretation or transcription.

7. DATA SYNTHESIS

The approach to synthesising included studies varies substantially, and some approaches are more effective at ensuring objectivity and minimising potential bias than others.

Is the choice of synthesis approach appropriate?

If appropriate, data should be pooled in a quantitative synthesis (e.g. meta-analysis, meta-regression). If substantial differences between populations, interventions, comparators or outcomes exist, meta-analysis (i.e. combining effect sizes across different studies) may not be appropriate. Since meta-analysis effectively treats all individual studies part of one large study, meta-analysis is only appropriate when calculating an average effect is meaningful. If it is not appropriate to pool data across studies in meta-analysis, a reason for this should be given, and structured approach to some other quantitative or narrative synthesis taken, with efforts made to make sense of the whole of the data set, beyond describing results from individual studies in turn, noting differences in the weight of evidence behind statements made, and appropriate use of table and graphical presentations of results. Vote-counting (summing the studies which gave positive or negative findings) is not an appropriate synthesis method as an indication of impact or effectiveness.

Is a statistical estimate of pooled effect (or similar) provided together with measure of variance and heterogeneity among studies?

If a sufficient quantity and quality of data is available then the presentation and assessment of evidence can be much improved by providing statistical information. Some evidence reviews will be unable to do this because of limitations of the primary data.

Is variability in the study findings investigated and discussed?

Studies can differ in their results (heterogeneity) which may be due to chance, but could also reflect variables other than the factor of interest that differ between studies (effect modifiers). The presence and magnitude of effect modifiers can reveal important information about a system. Investigating heterogeneity therefore indicates the degree to which effects are generalisable across taxa, regions etc., and is also necessary to evaluate the appropriateness of combining studies conducted on different populations or reporting different outcome metrics. These can be investigated statistically in meta-analysis through for example subgroup analyses, sensitivity analyses and meta-regression. In narrative syntheses differences in findings may be discussed in terms of differences in study design, context, population, focus etc.

8. LIMITATIONS

All reviews will have limitations and it is important that authors are explicit about the known limitations of the primary data and the conduct of the review process. Here we acknowledge the subjective nature of this criterion and the appraiser must use some subjective judgement to decide on the adequacy of any statement on limitations.