PRIFYSGOL
BANGOR
UNIVERSITY

Evidence maps and evidence gaps

O'Leary, Bethan; Woodcock, Paul; Kaiser, Michel; Pullin, Andrew

Environmental Evidence

1   **Title: Evidence maps and evidence gaps: evidence review mapping as a method for**

2   **collating and appraising evidence reviews to inform research and policy**

3

4   Bethan C. O'Leary[1*,] Paul Woodcock[1,2], Michel J. Kaiser[3], Andrew S. Pullin[1],

5

6   [1] Centre for Evidence-Based Conservation, School of Environment, Natural Resources

7   and Geography, Bangor University, Gwynedd LL57 2UW, UK

8   [2] Joint Nature Conservation Committee, Monkstone House, Peterborough, PE1 3JY,

9   UK

10   [2] School of Ocean Sciences, Bangor University, Menai Bridge, Anglesey, UK,

11   LL59 5AB

12   * Corresponding author

13   Author emails:

14   Bethan O'Leary: bethancoleary@gmail.com

15   Paul Woodcock: paul.woodcock@jncc.gov.uk

16   Michel Kaiser: michel.kaiser@bangor.ac.uk

17   Andrew Pullin: a.s.pullin@bangor.ac.uk

18

19

20

1

1 **Abstract**

2 Evidence reviews are a key mechanism for incorporating extensive, complex and

3 specialised evidence into policy and practice, and in guiding future research. However,

4 evidence reviews vary in scope and methodological rigour, creating several risks for

5 decision-makers: decisions may be informed by less reliable reviews; apparently

6 conflicting interpretations of evidence may obfuscate decisions; and low quality

7 reviews may create the perception that a topic has been adequately addressed, deterring

8 new syntheses (cryptic evidence gaps). We present a new approach, evidence review

9 mapping, designed to produce a visual representation and critical assessment of the

10 review landscape for a particular environmental topic or question. By systematically

11 selecting and describing the scope and rigour of each review, this helps guide non-

12 specialists to the most relevant and methodologically reliable reviews. The map can

13 also direct future research through the identification of evidence gaps (whether cryptic

14 or otherwise) and redundancy (multiple reviews on similar questions). We consider

15 evidence review mapping a complementary approach to systematic reviews and

16 systematic maps of primary literature and an important tool for facilitating evidence-

17 based decision-making and research efficiency.

18 **Keywords:** CEESAT; evidence-based policy; evidence review map; gap analysis;

19 review evaluation; research synthesis; research methods.

20

21

22

23

2

1   **Background**

2   Scientific evidence is central to effective environmental policymaking and practice but

3   its use requires an appreciation of the reliability of the evidence base. Primary research

4   forms the backbone of an evidence base; however, non-specialists may lack the

5   resources or expertise to evaluate the appropriateness of methodology and data analysis

6   in primary studies, and to identify trends and patterns across multiple studies.

7   Furthermore, the inherent complexity and variability of natural systems combined with

8   differences in study methods typically generates findings that can be selectively used to

9   support particular conclusions [1, 2]. Against this backdrop, non-specialists seeking an

10 overview of particular topics (e.g. decision-makers and researchers in other fields) are

11 increasingly likely to rely on evidence reviews that synthesise evidence across the

12 spectrum of primary literature related to a specific, policy-relevant question [3-6].

13 Evidence reviews (hereafter also referred to as 'reviews') attempt to answer a specific

14 question by aggregating and synthesising the results of primary studies and may include

15 meta-analysis (statistical methods for combining the magnitude of the outcomes [effect

16 sizes] across different data sets addressing the same research question [7]) and/or

17 narrative synthesis (use of prose to summarise and draw conclusions from primary

18 research which may be supplemented by the reviewers' own experience and may

19 include limited quantitative analysis [6]). Evidence reviews may or may not be

20 conducted within the framework of systematic review methodology [8]. Reviews that

21 only collect and configure the primary literature with respect to a broad question, such

22 as systematic maps, are not considered as an evidence review in this context.

23 High quality evidence reviews can provide decision-makers (or their advisors) with

24 quick and easy access to available evidence on a topic or question of interest. However,

25 as with primary research, reviews can differ in the rigour of the methods and the

3

reliability of findings [e.g. 6, 9, 10-12], and subtle differences in scope may influence their applicability to a particular problem. Indeed, the majority of reviews in environmental science are not conducted according to predefined guidelines, but instead apply a range of methods that each promote or compromise reliability to varying degrees [e.g. 6]. Differences in methodological reliability amongst reviews on similar topics create several related risks for researchers and decision-makers:

(i)   decisions may be informed by less rigorous and/or biased reviews because of a lack of systematic collation of reviews and subsequent appraisal of review quality;

(ii) apparently conflicting interpretations of evidence among reviews with similar scope may obfuscate decisions; and

(iii) no new or updated reviews are conducted on a topic, because researchers and decision-makers are unaware that the topic lacks a highly rigorous synthesis (cryptic evidence gaps [12]).

Clearly communicating the scope and reliability of an evidence-base to decision-makers and other end-users is therefore essential to ensure potential limitations in the conduct of the review(s) being considered are appreciated. However, in the absence of communication mechanisms tailor-made for use within decision-making processes this can be challenging. Indeed, difficulties in locating relevant evidence and assessing the reliability of information gathered are often amongst the main concerns highlighted by decision-makers [13].

Systematic review and systematic map methodologies were developed in part in recognition of the variable reliability of reviews [14, 15] and as an attempt to reduce these risks and provide high quality evidence synthesis and overviews for decision-

4

makers. Nonetheless, while systematic reviews are becoming more widespread in the environmental sector, not all conform to recognised standards [e.g. 16] and non-systematic evidence reviews still dominate the review landscape [17]. Moreover, in the medical sector, where systematic review terminology was coined and application of the methodology is most widespread [14, 15], the exponential rise in systematic reviews and meta-analyses have been dogged by criticisms that many do not follow full systematic review guidelines, are conflicted by pre-conceived opinions or financial motivations of the authors, and/or have been used to advance industry interests instead of good science [18]. With awareness and application of systematic review methodology expanding in the environmental sector, it is vital that ways to maintain and monitor methodological standards are developed and applied to ensure the objectivity, robustness and value of evidence reviews for decision-makers.

Tools for critically assessing the methodological reliability of individual reviews have developed in many sectors [see 19 and references therein], although limited techniques exist to meaningfully apply and integrate these assessment protocols to inform environmental policy. Similarly, in other sectors, methods to inform non-specialists on available evidence have developed [e.g. 20], but often focus on synthesising findings from the systematic review literature [21], or on describing and appraising systematic reviews and related impact evaluations [22, 23]. None are designed to describe and critically appraise the evidence review literature as a whole or explicitly consider studies that evaluate environmental outcomes arising from interventions. Accordingly, new methods to assess and communicate the reliability (including limitations of evidence and methodological rigour) and scope of all reviews, systematic and non-systematic, could represent a more viable alternative for summarising evidence reviews in the environmental sector. In response, we develop a method that we term 'evidence

5

review mapping', to produce a critical overview of all reviews examining the effectiveness of a given intervention and/or impacts of human pressures and management (e.g. effects of fisheries, impacts of land-use change, effectiveness of conservation interventions, etc.). This overview includes a systematic assessment of the questions addressed by each review (i.e. scope and relevance) combined with a critical appraisal of review methods (reliability and risk of bias). Outputs from evidence review mapping are designed specifically to inform non-specialists and improve communication of the evidence base by identifying the most relevant and reliable reviews, and to assist future syntheses by highlighting gaps and redundancy (multiple reviews on similar topics) in the review literature. Evidence review maps are tailor-made for the environmental decision-making community, offering a communication tool that consists of matrices that summarise the quantity and methodological rigour of reviews on a range of related questions, together with a series of supporting tables that provide more detailed information on the reviews for each particular question. Evidence review maps therefore do not aim to answer a specific question but rather intend to enable end-users to quickly assess the volume of evidence on the question(s) of interest, and to obtain an overview of how reliable that evidence is.

Here, we describe how to construct evidence review maps to inform environmental policy and research, providing examples with reference to a study we undertook in conjunction with developing this methodology [12]. We propose that evidence review mapping offers a complementary approach to systematic reviews and systematic maps, and suggest that adoption of the methodology will facilitate evidence-based policy and practice in conservation and environmental science.

Methodology - Evidence review mapping

Our approach to evidence review mapping consists of the following steps: (1) define the overall question of interest, construct a series of more refined questions that consider key aspects of the overall question, and then design the search strategy; (2) systematically search and screen for relevant evidence reviews; (3) assess the scope of each evidence review against the questions defined in step 1; (4) critically appraise the methods of each evidence review using a standardised protocol (we use the Collaboration for Environmental Evidence Synthesis Assessment Tool – CEESAT [19]); and (5) construct the evidence review map(s). The production of an evidence review map integrates some core systematic review methods complemented by several novel approaches designed specifically to search for, collate, categorise, and communicate review articles. We provide a description of each stage below illustrated, where appropriate, with details adapted from Woodcock et al. [12] who examined the evidence review landscape for the question 'What is the effectiveness of marine protected areas as a tool for mitigating the impacts of fisheries on biodiversity?'

1. Define the question of interest, construct a series of more refined questions, and then design the search strategy

To provide the framework for the evidence review map, the overall question of interest (i.e. scope of the map) should be established. As with systematic review methodology the question will largely determine the inclusion criteria for the reviews that form the evidence review map. Consequently, we recommend that a Population, Intervention/Exposure, Comparator, Outcome (PI/ECO) structure is used to ensure a clearly defined question is developed [8].

Evidence review mapping uses a hierarchy of questions to assess the scope of reviews: the overall question sets the scope of the map and a series of more refined questions

1 then establish the key areas of interest within the overall question. More refined

2 questions may be separated into broad questions (considering one key area of interest)

3 and specific questions (considering paired combinations of key areas, based on the

4 Population, the Intervention/Exposure, and the Outcome Metrics). Figure 1 describes

5 the question hierarchy framework illustrated with example questions from Woodcock et

6 al. [12]. Compiling a comprehensive list of key areas of interest and their pairwise

7 combinations generates the framework for the evidence review map (Figure 2). The

8 number of refined questions represents a pragmatic trade-off between capturing the

9 many potential influences on the overall question of interest and generating an

10 unmanageably large number of specific questions. The scope of the evidence review

11 map will therefore partly depend on the resources available. In addition, while

12 researchers may undertake evidence review mapping independently to guide future

13 studies, maps that are designed to inform policy should develop the evidence review

14 map framework in consultation with stakeholders in order to ensure relevance through

15 appropriate selection of the Population, Intervention/Exposure, and Outcome Metrics to

16 consider.

17 Together with the question, it is important to explicitly document the criteria for

18 deciding on whether or not articles are relevant to include to ensure objectivity,

19 transparency and repeatability during article screening. Once these have been defined,

20 an appropriate search strategy should be developed and detailed within an *a-priori*

21 protocol. The search strategy should draw on search methods used for systematic

22 reviews [8] with the search effort depending on the scale of the evidence review map,

23 the volume of subject-specific evidence, and the resources available. Topic-specific

24 search strings should then be narrowed to focus on review articles using terms such as

25 'AND (review OR "meta-analy*" OR synthes*)' or using appropriate database filters

for 'review articles' if available and known to be reliable, and the databases that will be searched should be documented [see 8 for further information on search strategy design and reporting]. Note that in sectors where systematic reviews are more widespread, search filters designed to retrieve research by study design or focus have been heavily invested in (e.g. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home). While similar filters exist within search engines commonly used by the environmental sector (e.g. Web of Science, Scopus) this functionality is less well-developed and database-specific so caution is recommended before relying solely on their use.

2. Systematically search and screen for relevant evidence reviews

A systematic search and screening process to identify relevant articles should be undertaken in line with systematic review guidelines [8], with searches comprehensively documented and the repeatability of inclusion decisions during screening tested using a kappa test of agreement [24, 25] or similar. Inclusion criteria should be refined as necessary to ensure repeatability [8]. Articles assessed for relevance at full text should be clearly documented in the supporting tables to the map (see step 5) with the reasons for exclusion provided where appropriate to maintain transparency. Importantly, in some instances, reviews that are excluded may contain related information of potential interest (e.g. outside the taxonomic or geographic scope of the evidence review map), or partially consider some questions of interest as part of a broader-ranging review with a different scope (this can occur particularly for narrative reviews). Note also that some studies use meta-analytical techniques in the analysis of long-term primary data [e.g. 26] or data from selected case studies [e.g. 27] rather than with the aim of comprehensively synthesising published research. As these would not be expected to follow all of the methods required to produce a rigorous review of

1 primary research (e.g. *a-priori* protocol, comprehensive searching, screening), they are

2 unsuitable for evaluation using CEESAT. However, marking such reviews as

3 'borderline relevant' can assist decision-makers seeking additional information.

4 Included articles can be assigned a number (meta-analyses) or letter (narrative

5 syntheses) to act as a unique identifier when constructing the evidence review map and

6 these should be documented in supporting tables (e.g. Table 1a).

7 3. Assess evidence review scope

8 Constructing an evidence review map requires that each relevant review is

9 systematically categorised according to the question(s) addressed (as defined in step 1)

10 and the type of synthesis undertaken (e.g. narrative/qualitative [which may include

11 limited quantitative analyses] or meta-analyses; see [6] for definitions of each). Note

12 that multiple questions are often addressed within a single review, and so a single

13 review may be included several times in an evidence review map. The extent to which

14 scope can be objectively categorised is influenced by the methods employed in the

15 review. Whilst a meta-analysis can usually be objectively categorised as addressing a

16 particular question based on whether or not effect sizes are presented, there is no such

17 obvious distinction in many narrative reviews, in which questions could be addressed

18 through varying amounts of text with varying degrees of relevance and supporting

19 references. This problem is exacerbated because the scope of narrative reviews is often

20 broader than meta-analytical reviews. Reliable categorisation of scope is thus possible

21 in greater detail for meta-analyses than for narrative syntheses. The assessment of

22 review scope should therefore be undertaken in two parts, firstly considering reviews

23 that apply meta-analytical techniques and secondly reviews that use narrative synthesis.

Categorisation of meta-analyses as addressing particular questions requires effect sizes to be quoted directly, presented graphically or used in statistical tests of relationships [12]. Instances where relevant terms are included as potential confounding variables but statistics (e.g. effect sizes) are not reported would not be considered as directly addressing a given question [12]. A threshold for the minimum number of primary studies a meta-analysis must contain to be categorised as addressing a particular question could be set. The minimum threshold is highly context-specific (e.g. relating to the quality of primary research, typical effect sizes and variances, etc.) and consequently requires a transparent case-by-case judgement for each evidence review map. Where a threshold is considered appropriate, reviews that do not meet this threshold should be noted as partially addressing the question, thereby allowing articles that are based on a small volume of primary research to be identified.

Categorisation of narrative syntheses should be undertaken wherever possible using the refined questions initially devised in step 1. However, because narrative syntheses often cover a range of topics in varying depths, such fine-scale categorisation may not be possible and so questions may need to be broadened to more accurately reflect narrative review content (e.g. by broad area of focus etc. see Figure 3 and [12]).

4. Critically appraise the methods of each evidence review using a standardised protocol

The assessment of review methodology forms the penultimate stage in evidence review map development. A standardised protocol designed to assess the reliability of environmental evidence reviews should be utilised to critically appraise the methodological rigour of each relevant review in a consistent manner. For this purpose we recommend the Collaboration for Environmental Evidence Synthesis Assessment

11

1 Tool [CEESAT; 19]. The current version of CEESAT (available at

2 www.environmentalevidence.org/review-appraisals) consists of 13 criteria relating to

3 the reliability (combining objectivity, transparency, and comprehensiveness) of reviews

4 [see 19 for details], and achieves good repeatability when independent assessments of

5 the same review are compared [6, 12, 19]. For each criterion, reviews receive 3 points,

6 1 point, or 0 points. Scores therefore range from 0-39: the higher the score, the greater

7 the confidence that the review methodology is robust and reliable in terms of

8 repeatability and risk of bias. Importantly, while certain criteria within CEESAT require

9 statistical analysis to score highly, points for these criteria are available to narrative

10 syntheses [6, 19]. Furthermore, high scores are available equally to narrative syntheses

11 and meta-analyses for most of the criteria enabling narrative syntheses to be assessed as

12 having high reliability (a score of 26.5+) where appropriate [e.g. see 6]. Reviews should

13 be independently scored by two assessors and the repeatability of the assessment

14 evaluated with a weighted kappa test of agreement [24, 25] or similar to take into

15 account the magnitude of any disagreements, e.g. a 1-0 disagreement is ranked as

16 magnitude 1, whereas a 3-0 disagreement is ranked as magnitude 3 [10, 28].

17 Disagreements between assessors should be discussed and where these reflect

18 uncertainty over whether or not a criterion was met, the average score from the two

19 assessors should be used.

20 There are a number of possible approaches to interpreting CEESAT scores [see 19 for

21 further discussion] however, we currently recommend dividing total CEESAT scores

22 into three categories 0-13, 13.5-26 and 26.5+, loosely representing low,

23 intermediate/moderate and high methodological reliability. The boundaries for these

24 categories reflect an average score across the 13 criteria of 0-1, 1-2 and 2-3. Note that

25 while these boundaries may change as further guidance on scoring interpretation

12

becomes available, or if certain aspects of review conduct are prioritised by those

conducting an evidence review map, the methodology for incorporating scores into

evidence review mapping will remain valid.

5. Construct the evidence review map

Finally, using information from steps 3 and 4, a series of evidence review maps may be

constructed to visually represent the review landscape for the overall question of

interest (Figure 3). Separate maps should be constructed to describe meta-analyses and

narrative reviews to ensure similar levels of objectivity in review categorisation within

each map. Evidence review maps should be constructed using refined questions as

defined in step 1 for meta-analyses and those determined in steps 1 and/or 3 for

narrative reviews. Note that for the example here, none of the narrative syntheses

provided sufficient information to score highly when assessed with CEESAT [e.g. see

12] and, as a consequence, the narrative evidence review map shows all reviews to be

of low reliability (Figure 3b). This reflects the specific evidence base for MPA

effectiveness rather than being a consequence of differences in the way in which

CEESAT evaluates narrative syntheses *vs*. meta-analyses [12].

Evidence review maps consist of a matrix that combines information on the number of

reviews addressing a given question and the methodological rigour of each review,

enabling end-users to see what evidence there is on the question(s) they are interested

in. The matrix overview is supported by a series of tables that allow the most rigorous

reviews on each question to be identified. The matrix should be read using

combinations from the top and left headings to form a particular question. Doughnut pie

charts can be created to represent (1) the total number of reviews that address each

individual question (included in the centre of the doughnut pie) and (2) the proportion

13

of those reviews that are of high, medium or low methodological reliability. Symbols should be used to identify where reviews have been categorised as partially addressing a particular question (due to the threshold for number of included primary research articles not being met). The format of the matrix means that some questions will not be applicable; these areas should be left blank. Full details of reviews included for each specific question, together with details of any reviews that partially address a given question should then be provided in a series of supporting tables (e.g. see Table 1b and c and [12]).

Supporting tables should include: (1) details for the search strategy; (2) a list of relevant reviews with their unique identifier and review score; (3) a list of excluded studies with reasons for exclusion; and (4) a series of tables detailing the meta-analyses and narrative syntheses examining each refined question, designed to direct end-users to the most relevant and rigorous review for their requirements.

Discussion

Understanding the reliability of an evidence base is central to effective decision-making and developing mechanisms for communicating this to decision-makers is therefore essential. While systematic review methodology is considered a key tool for unbiased evidence synthesis, the reliability of evidence reviews will continue to vary for many reasons [18]. With the number of reviews of all types continuing to increase, evidence review maps provide the opportunity to visualise the review landscape for an overall question of interest and to guide non-specialists to more relevant and reliable reviews. We consider evidence review mapping a complementary approach to systematic reviews and systematic maps and an important tool for facilitating evidence-based decision-making.

14

Evidence review mapping relies on systematic searching, transparent decisions on article inclusion and exclusion, objective assessment of review scope and a standardised and repeatable protocol for critically appraising individual reviews. Application of our approach has illustrated the variable scope and reliability of published evidence reviews and the need to ensure non-specialists can locate the most relevant and rigorous reviews on particular questions of interest, as well as indicating how planned reviews can be designed to complement the existing body of reviews [12]. We believe this approach and its outputs will be useful to decision-makers, advisors and knowledge brokers wishing to use evidence in environmental policy and practice, as well as to researchers looking to contribute to the evidence base through targeted evidence synthesis. Our approach to evidence review mapping could be applied widely to many important questions in environmental policy, as an 'evidence service' with considerable benefits for research efficiency and evidence-based policy.

Considerations for conducting evidence review maps

While evidence review mapping is a valuable tool, it will pose some challenges to those wishing to construct such maps. Most notably, decisions over whether or not reviews are relevant for inclusion require subjective judgement. This difficulty arises particularly in narrative syntheses, because there is a continuum between studies that exclusively review the findings from relevant primary research versus studies that have a very broad scope or a more conceptual focus and are therefore less appropriate for evidence review mapping. Because most subjective decisions on relevance relate to narrative reviews, altering these decisions would not affect the meta-analytical evidence review map and, while they might adjust the average narrative review score for a given question, they are unlikely to markedly change the conclusions on review rigour and scope. Nonetheless, ensuring transparency of decisions at all stages of evidence review

15

mapping by documentation in the supporting tables is an important component of the methodology to enable end-users to understand and challenge the decisions made over article inclusion and categorisation. Additionally, some methodologically distinct forms of review, such as qualitative syntheses or mixed methods may not be suitable for appraisal using CEESAT and including such reviews will require further research and development.

Evidence review maps rely on the use of a standardised scoring tool to assess the reliability of reviews. Like other scoring tools, CEESAT assesses the likelihood that a review is reliable on the basis of key attributes relating to available evidence, conduct and reporting standards, and does not guarantee the reliability of a review against other factors such as author errors. In addition, total scores of reviews can mask specific strengths or weaknesses across criteria. A breakdown of scoring across individual criteria may therefore be a useful subsequent output to ensure that decision-makers can gauge the extent to which the strengths and weaknesses of a review make it suitable for the intended use [see 19 for a detailed discussion of important caveats in applying and interpreting CEESAT scores]. Researchers who wish to undertake evidence review mapping may wish to use alternative boundaries and/or weightings of criteria to represent reliability than those suggested here if certain aspects of review conduct are viewed as particularly important to the end-user. In such instances, clear rationale for amending these boundaries and/or weightings should be provided as part of the evidence review map.

There may be instances in which more than one high scoring review addresses a particular question. In such situations, further assessment could consider the consistency in findings between reviews (noting that direct comparisons of specific results can be misleading if subtle differences in review scope are not identified). If

results differ between reviews, potential reasons for ambiguity could then be considered, and further work targeted to examine the evidence base where reasons for discrepancy are unclear. In the latter situation, systematic reviews, containing meta-analytical techniques wherever possible, and/or targeted and well-designed primary research are recommended to ensure that policymaking is informed by reliable evidence that is robust and methodologically rigorous [29].

Finally, note that unlike 'review of reviews' that aim to provide a synthesis of evidence from more than one review, evidence review maps do not set out to answer a specific question but rather seek to provide an overview of the existing review evidence base. Consequently, maps are intended to guide decision-makers to relevant information and illustrate strengths and weaknesses in the evidence base, rather than to directly provide policy recommendations or guidelines. Future work that may add value to evidence review maps might include developing user-friendly summaries on included reviews, or reports summarising the findings of the evidence review map together with implications for policy and research.

Conclusions

As the review literature continues to expand, it will become increasingly difficult for non-specialists to locate all relevant evidence reviews. Furthermore, when selecting reviews to inform decision-making, non-specialists may lack the resources to critically appraise all available syntheses and may instead treat all evidence reviews equally, or use measures of review rigour that are questionable and/or subjective (e.g. journal impact factor, citation count, author reputation). However, we, and others [e.g. 6, 9, 12, 30], have found that published evidence reviews in the environmental sector vary considerably in reliability and scope, which presents challenges to those wishing to

17

undertake evidence-based decision-making. We therefore propose that evidence review

mapping represents an important method for communicating the reliability and scope of

all reviews on a particular topic to non-specialists, thereby facilitating evidence-based

policy and practice in conservation and environmental science.

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Competing interests: The authors declare that they have no competing interests"

Authors' contributions: Led the research: BCO, PW, ASP. Wrote and reviewed the

manuscript: BCO, PW, MJK, ASP. All authors read and approved the final manuscript.

References

1. Egger M, Smith GD. Bias in location and selection of studies. BMJ. 1998;316:61-66.

2. Schott GD. The reference: more than a buttress of the scientific edifice. J R Soc Med. 2003;96:191-193.

3. Seavy NE, Howell CA. How can we improve information delivery to support conservation and restoration decisions? Biodivers Conserv. 2010;19:1261-1267.

4. Cook CN, Carter RW, Fuller RA, Hockings M. Managers consider multiple lines of evidence important for biodiversity management decisions. J Environ Manage. 2012;113:341-346.

5. Pullin AS, Knight AT, Stone DA, Charman K. Do conservation managers use scientific evidence to support their decision-making? Biol Cons. 2004;119:245-252.

6. O'Leary BC, Kvist K, Bayliss HR, Derroire G, Healey JR, Hughes K, et al. The reliability of evidence review methodology in environmental science and conservation. Env Sci & Pol. 2016;64:75-82.

7. Koricheva J, Gurevitch J, Mengersen K, *Handbook of Meta-Analysis in Ecology and Evolution*. 2013: Princeton University Press.

8. CEE, *Guidelines for systematic review and evidence synthesis in environmental management*. 2013, Version 4.2. Environmental Evidence.

9. Roberts PD, Stewart GB, Pullin AS. Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. Biol Cons. 2006;

1   10.   Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External
2   validation of a measurement tool to assess systematic reviews. PLoS ONE.
3   2007;2:e1350.

4   11.   Philibert A, Loyce C, Makowski D. Assessment of the quality of meta-analysis in
5   agronomy. Agricult Ecosys Environ. 2012;148:72-82.

6   12.   Woodcock P, O'Leary BC, Kaiser MJ, Pullin AS. Your evidence or mine? Systematic
7   evaluation of reviews of marine protected area effectiveness. Fish Fish. in press;

8   13.   Holmes J, Clark R. Enhancing the use of science in environmental policy-making and
9   regulation. Environmental Science & Policy. 2008;11:702-711.

10   14.   Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for
11   clinical decisions. Ann Intern Med. 1997;126:376-380.

12   15.   Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. Eval Health
13   Prof. 2002;25:12-37.

14   16.   O'Leary BC, Bayliss HR, Haddaway NR. Beyond PRISMA: Systematic reviews to inform
15   marine science and policy. Mar Pol. 2015;62:261-263.

16   17.   Haddaway NR, Woodcock P, Macura B, Collins A. Making literature reviews more
17   reliable through application of lessons from systematic reviews. Conserv Biol. 2015;

18   18.   Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted
19   Systematic Reviews and Meta-analyses. Milbank Q. 2016;94:485-514.

20   19.   Woodcock P, Pullin AS, Kaiser MJ. Evaluating and improving the reliability of evidence
21   syntheses in conservation and environmental science: a methodology. Biol Cons.
22   2014;176:54-62.

23   20.   Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A
24   systematic review of published evidence maps and their definitions, methods and
25   products. Systematic Reviews. 2016;5:28.

20

21. Caird J, Sutcliffe K, Kwan I, Dickson K, Thomas J. Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach. Evid Policy. 2015;11:81-87.

22. Snilstveit B, Vojtkova M, Bhavsar A, Gaarder M, *Evidence Gap Maps - A tool for promoting evidence-informed policy and prioritizing future research*. 2013, World Bank Policy Research Working Paper No. 6725.

23. McKinnon MC, Cheng SH, Garside R, Masuda YJ, Miller DC. Sustainability: Map the evidence. Nature. 2015;528:185-187.

24. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46.

25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1997;33:159-174.

26. Ojeda-Martinez C, Bayle-Sempere JT, Sanchez-Jerez P, Forcada A, Valle C. Detecting conservation benefits in spatially protected fish populations with meta-analysis of long term monitoring data. Mar Biol. 2007;151:1153-1161.

27. Vandeperre F, Higgins RM, Sanchez-Meca J, Maynou F, Goni R, Martin-Sosa P, et al. Effects of no-take area size and age of marine protected areas on fisheries yields: a meta-analytical approach. Fish and Fisheries. 2011;12:412-426.

28. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37:360-363.

29. Pullin AS, Knight TM. Doing more good than harm - building an evidence-base for conservation and environmental management. Biol Cons. 2009;142:931-934.

30. Huntington BE. Confronting publication bias in marine reserve meta-analyses. Front Ecol Environ. 2011;9:375-376.

1   **Figure Legends**

2   **Figure 1.** Question hierarchy framework used to assess the scope of reviews for

3   evidence review mapping. Descriptions of the purpose of each question level are

4   provided to the left of the triangle. Example questions based on Woodcock et al. [12]

5   illustrating each level of questioning are provided to the right of the triangle.

6   **Figure 2**. Schematic illustrating the process of constructing the framework for an

7   evidence review map. Example evidence review map explores the effectiveness of

8   marine protected areas (MPAs) for biodiversity conservation. Key components of the

9   overall question are identified in the left panel (e.g. Population: regional and taxonomic

10   focus, Intervention: aspects of MPA design considered, Outcome: outcome metrics used

11   to assess MPA effectiveness). In the panel on the right, these components are combined

12   to construct the framework for the evidence review map consisting of broad (dark grey

13   boxes) and specific (light grey boxes) questions. White boxes indicate questions that are

14   not applicable, e.g. Global/Temperate question combinations. Abbreviations in

15   headings refer to: Taxa − Invert=invertebrates, MPA Char=MPA Characteristics,

16   Outcome Measures - Abund=abundance. Example adapted from Woodcock et al. [12].

17   **Figure 3**. Example (a) meta-analytical and (b) narrative evidence review map

18   illustrating the marine protected area review landscape. The matrix should be read using

19   combinations from the top and left headings to form the particular question of interest.

20   Each individual doughnut chart describes the number of reviews addressing a question,

21   and the proportion of reviews that are high (26-5+; black), moderate (13.5-26; grey) and

22   low (≤13; white) scoring. Star symbols represent where one or more reviews have been

23   identified that partially address the particular question. Full details identifying reviews

24   that address a particular question are reported in the supporting tables. Abbreviations in

22

1 headings refer to: Taxa – Invert=invertebrates MPA Char=MPA Characteristics,

2 Outcome Measures - Abund=abundance, Broad Focus – BD-biodiversity,

3 Fish=Fisheries. Example adapted from Woodcock et al. [12].

4

5 **Tables**

6 **Table 1.** Example evidence review map supporting tables (a) reviews assessed as

7 relevant for inclusion, with review score, (b) scope of meta-analyses that examine broad

8 questions: region, taxa, MPA characteristic and outcome measure, and (c) scope of

9 narrative syntheses that examine the specific question: broad focus and region.

10 Superscript Roman numerals adjacent to 'Scope' in (b) are used to refer the end-user to

11 relevant notes on additional reviews that consider the question but with less than the

12 required number of primary studies or without reporting effect sizes. Example adapted

13 from Woodcock et al. [12].

14

a)

Review Number	Reference	CEESAT Score
1	Meta-analytical reference	
a	Narrative reference	

15

b)

	Scope	Review number
Region	Global	3, 6-9, 11-14, 16, 17
	Tropical [i]	8, 10
Taxa	Fish [ii, iii, iv]	1-3, 5, 6, 8, 10-18
	Invertebrate [ii, iii, iv]	6, 8, 15, 16, 18
MPA Characteristic	Size [i, vi]	1-3, 5, 6, 8, 10, 11, 15, 17, 18
	Age [v]	1-3, 5, 7, 10-13, 15, 17, 18
Outcome Measure	Abundance	1-18
	Biomass [vii, viii, ix]	4-8, 17, 18

c)

	Scope	FOCUS	
		Biodiversity	Fisheries
Region	Global	b, c, f, h, i	b, e, f, g, h, i
	Temperate	d	d
	Tropical	a	

16

17

18