Assessing the evidence of ecosystem services studies: a framework and its application

Anne-Christine Mupepele [1,2] & Carsten F. Dormann[1]

[1]University of Freiburg

[2]anne-christine.mupepele@biom.uni-freiburg.de

Abstract

Reliability of scientific findings is always important, but even more so in ecosystem service science, where results can directly lead to policy implementations. The assessment of reliability resulted in the introduction of the evidence-based concept in the medical sciences. It aims at identifying the best available information to answer the question of interest. In environmental sciences, the evidence-based concept is well developed [very few people actually use it!] in conservation, but so far no guidelines exists for ecosystem services science. To apply the evidence-based concept in practice we need a scale to rank study designs commonly used in ecosystem services science as well as an assessment of the actual quality of implementation in a specific study. Here, we define a framework for evidence-based ecosystem service science, together with a design scale and a critical-appraisal checklist. The approach is described with detailed examples and specific suggestions for different user groups.

Keywords: conservation - governance - level-of-evidence hierarchy - management - quality checklist - quantification - rigour - valuation

Ecosystem services, the benefits humans derive from nature (Millennium Ecosystem Assessment, 2005), have become a central paradigm for environmental science, management and legislation (Daily *et al.*, 2009). Being at the interface of policy and ecology, they straddle the gap between scientists and decision-makers (Raffaelli and White, 2013). Ecosystem services investigations may result directly into actions and it is therefore crucial to assess the reliability of current ecosystem services investigations and recommendations (Boyd, 2013).

Assessing the reliability of scientific statements has been translated into the evidence-based concept (Sackett *et al.*, 1996; GRADE Working Group, 2004; OCEBM Levels of Evidence Working Group, 2011, Cochrane Collaboration - www.cochrane.org). The concept aims at identifying the best available information to answer the question of interest (Sackett *et al.*, 1996). Pioneered in medicine, the evidence-based concept today is used by several other disciplines, e.g. justice (www.campbellcollaboration.org), conservation science (Pullin and Knight, 2001, 2009; Sutherland *et al.*, 2004) or forestry (Binkley and Menyailo, 2005; Petrokofsky *et al.*, 2011).

Well over ten years ago, Sutherland (2000) and Pullin and Knight (2001) introduced the evidence-based concept to conservation. Today, the Collaboration for Environmental Evidence (www.environmentalevidence.org) fosters the creation of systematic reviews to reach highest possible evidence (Bowler *et al.*, 2012; Collaboration for Environmental Evidence, 2013, see also *Journal for Environmental Evidence*). Communicating the evidence is an important next step, with focus on applicability and user friendliness for decision makers (Dicks *et al.*, 2014, *Journal for Conservation Evidence*). Traditional biodiversity conservation approaches and the ecosystem services concept have their overlaps (Nelson *et al.*, 2009), but currently the evidence-based concept is not used in ecosystem services science. This may in part be due to ignorance on how to incorporate all aspects of ecosystem services (e.g. valuation) in the evidence-based concept. Further more, a clear hierarchical ranking for study designs used in ecosystem services is not yet available.

In this contribution we provide guidance on the terminology around evidence-based practice, to ensure that scientists and practitioner can communicate effectively across the disciplines and backgrounds. Next, we detail how to use the evidence-based concept in ecosystem services, together with a new level-of-evidence framework ranking scientific rigour in ecosystem services science. Finally we illustrate the application of the concept with specific examples, and offer suggestions for the next steps.

The evidence-based concept

The terminology used around evidence-based practice is diverse and not always consistently used. However, a well-defined terminology is essential for effective communication between practitioners and scientists. According to the Oxford Dictionaries, **evidence** is the available body of facts or information indicating whether a belief or proposition is true or valid (www.oxforddictionaries.com/definition/english/evidence). In other words, evidence is a measure for the knowledge behind a statement (see also Higgs and Jones 2000, p.311). The **strength of evidence** reflects the quality of our knowledge and we can identify whether a statement is based on **high or low evidence**, hence very reliable or hardly reliable. **Evidence-based** practice implies that actions are undertaken based on the highest reliability available. It further means that if high evidence is missing, the end-user is aware of the low reliability of the recommendations for action.

Evidence-based practice starts with a question or a purpose (Fig. 1). This question is answered in research studies that follow a resolute study design. The study design is the set-up of the investigation, e.g. case control or observational design. These study designs are not equally compelling with respect to inferring causality. The differences in study designs translate into different strengths of evidence, and in order to identify the level of evidence, study designs are ranked hierarchically according to a level-of-evidence scale (Fig. 2). After identifying the underlying study design, also the implementation of the design must be thoroughly assessed in a critical appraisal. It evaluates quality aspects, such as sample size and randomization. A comprehensive list of quality aspects is provided by a quality checklist (Table 2), facilitating the critical appraisal. The combination of study design and quality criteria will allow the identification of the final level of evidence (Fig. 1).

1. Question, outcome and the context

The first step of evidence-based practice is to clearly identify the purpose of the investigation, ideally in the form of a question (Pullin *et al.*, 2009; Higgins and Green, 2011; Collaboration for Environmental Evidence, 2013, p.20-23). For ecosystem service science, we additionally require the specification of the environment. The information *which* ecosystem service is investigated in *which* system is necessary to determine the context for the validity of the answer.

Ecosystem service science is interdisciplinary and combines ecology, economy, political and other social and natural sciences. In order to know which field we operate in, we recommend to determine the focus of the ecosystem service question, either 'quantification', 'valuation', 'management' or 'governance' (Table 1). Ideally these foci are investigated in this order, starting with the quantification of an ecosystem service, which should then be valued; the most valuable services will be trialed by a well-adapted management option and in the end a governance strategy steering the preferred type of management is implemented. Deviations of this structure are common, e.g. valuation does not necessarily require prior quantification. However, when management or governance interventions are to be tested, evidence-based quantification or valuation is required to be able to measure the effect of the management or governance intervention (e.g. Acuña *et al.* (2013) uses valuation methods to determine success or failure of a management strategy).

So far we have determined the question and embedded it in its context and disciplinary focus. In an ecosystem services study, this is followed by the actual investigation. The outcome is usually the result of the study, the answer to the originally formulated question (Fig. 1).

2. Evidence assessment

The evidence assessment investigates study design and quality to determine the evidence of the outcome.

2a. Level-of-evidence scale

At the heart of evidence-based practice lies a hierarchy which ranks study designs. This level-of-evidence scale determines whether the investigation of interest yields high or low evidence (Fig. 2). **Systematic reviews (LoE1a)** are at the top end of the level-of-evidence scale and provide the most reliable information. They summarize all information gained in several individual studies, have an *a priori* protocol on design and procedure, and are conducted according to strict guidelines (e.g. Collaboration for Environmental Evidence, 2013). Ideally they include quantitative measures, at best a meta-analysis (in the strict sense; see Vetter *et al.*, 2013). Other more **conventional reviews (LoE1b)** may also include quantitative analysis or be purely qualitative. They both summarize the findings of several studies, but systematic reviews assess completeness and reproducibility more carefully and try to avoid publication bias by including grey literature (Higgins and Green, 2011).

The necessary condition for any review is that appropriate individual studies are available. The most reliable individual studies are **studies with a reference (LoE2)**. Typically, these are case-control or before-after control-impact studies **(LoE2a)**. Another useful approach can be the comparison of different methods, e.g. for the valuation of ecosystem services, where no 'true' reference exists. However the results between both methods have to be consistent to provide high evidence **(LoE2b)**.

Observational studies (LoE3) are individual studies without control. These include studies employing correlative or inferential statistics, e.g. testing for the influence of environmental variables on the quantity of an ecosystem service **(LoE3a)**. Descriptive studies imply data collection and representation without statistical testing (e.g. data summaries, ordinations, histograms, charts). In ecosystem service science these are often survey or expert elicitations **(LoE3b)**.

The lowest level of evidence are statements **without underlying data (LoE4)**. These are usually expert opinions, often not distinguishable from randomness (Tetlock, 2005). Other statements without underlying data are reasoning based on mechanism and 'first principles': *A* works according to a certain mechanism, so we expect *B* to work in the same way. These first principles are not reliable in ecology (Lawton, 1999).

It is important to note that 'method' and 'design' should not be confused. Methods are the means used to collect or analyse data, e.g. remote sensing, questionnaires, ordination techniques, model types. Design resents how the study was planed and conducted, e.g. a case-control or descriptive design. The same methods can be employed for different underlying designs. Remote sensing for example can be done purely descriptively or with a reference such as ground-truthing or in a 'before-after' design.

2b. Critical appraisal

The critical appraisal assesses the implementation of a study design. A study with a high-evidence design may be poorly conducted. The critical appraisal identifies the study and reporting quality and may lead to a downgrading of the level of evidence. This depends on objective, sometimes design- (e.g. review) or focus- (e.g. valuation) specific criteria. Several publications provide lists with quality criteria (e.g. Pullin and Knight, 2003; GRADE Working Group, 2004; Söderqvist and Soutukorva, 2006; Bilotta *et al.*, 2014, Oxford Centre for Evidence-Based Medicine 2011 www.cebm.net). We combined these lists to a general quality checklist (Table 2). The checklist consists of 44 questions with the possibility to use only a subset if some questions are not appropriate for the specific context. All questions answered with 'yes' receive one point (or two points if it is an important questions - in bold font in Table 2). In case of non-reported issues, we advice the answer 'no' to indicate a deficient reporting quality. The percentage of points received out of possible points will help to decide whether to downgrade the level of evidence (Table 3).

Reviews provide information on the highest level of evidence and their critical appraisal is different from other designs, because they themselves are, by definition, based on studies with lower evidence (see Table 2: section review). If only studies based on low evidence (e.g. LoE4) were included, then the review should be downgraded, regardless of other quality criteria.

We encourage the use of the checklist for an orientation, but we want to emphasise that this procedure can not be fully standardised. Quality aspects can also depend on the context of the study and the final judgement will remain with the user.

Application of the evidence-based concept

Among the most common application of the evidence-based concept is the **systematic review**. It summarises all knowledge available for a specific question. A systematic review is time consuming and if a specific answer is needed in a shorter time, a 'rapid evidence assessment' (UK Civilservice, 2013) may be a useful alternative. Another approach to evidence-based practice are **synopses**. Synopses do not focus on a specific question but bring together information from a much broader topic, e.g. from a whole animal group, such as bees (Dicks *et al.*, 2010). They assess the evidence of various possible interventions. A third possibility applying the evidence-based concept are **guidelines**. These 'best practice guides' recommend, based on best available evidence, methods to assess ecosystem services quantification, valuation, management and governance (see also Graham *et al.*, 2011). Guidelines are also created by answering questions. In comparison to systematic reviews, these questions place more emphasis on the methodological side, asking 'Which is the best method to measure CO_2 stored in temperate forests?' rather than 'How much carbon is stored in temperate forests?' as in a systematic review. This allows forest scientists to employ the best method to quantify carbon in any temperate forest. In the case of evidence-based ecosystem service science, guidelines would also identify the evidence base of common instruments and tools such as InVEST (Tallis and Polasky 2009, also compare with ‘summaries’ in Dicks *et al.* 2014).

All these possibilities for the application of the evidence-based concept summarise individual studies and therefore require the evaluation of the evidence of individual studies included. In systematic reviews this is typically done as one step of the critical appraisal. So far a specification was missing, how exactly this should be done. The instruments introduced here provide a clear guideline with an appropriate scale and a quality checklist to rank individual studies.

Examples of evidence-based practice

With the method described above we can assess the level of evidence of individual studies. In the following we provide several examples chosen from the literature (more details in the supporting information Table S1 and S2). The first example tackles the question: ‘How does adding dead wood influence the provision of ecosystem services?’ (Acuña *et al.*, 2013). The authors investigated two ecosystem services (fish and retention of organic and inorganic matter) in a river-forest ecosystem in Spain and Portugal and studied the effect of a management intervention. Their study design followed a before-after control-impact approach, which is LoE2a. The critical appraisal (see Table S2) revealed shortcomings: only 19 out of 29 points (66%) were achieved. The level of evidence was downgraded by one level to LoE3a. We therefore conclude that the statement made by Acuña *et al.* (2013): ‘restoration of natural wood loading in streams increases the ecosystem service provision’ is based on LoE3a.

A second example is the governance-related question by Entenmann and Schmitt (2013): 'Do stakeholders relate REDD+ to biodiversity conservation?' They found that synergies between REDD+ and biodiversity conservation were assumed by stakeholders. It is an observational, descriptive design (LoE3b), receiving only 11 of 23 quality points and therefore downgraded to LoE4.

The third example was a management-related systematic review of Bowler *et al.* (2010), conducted according to the guidelines of the Collaboration for Environmental Evidence (2013). They investigated the effect of greening urban areas on the air temperature to mitigate heat exposure. They found that green space in an urban area is on average 1°C cooler, than a built-up site. According to the quality assessment the study achieved 30 out of 31 points (97%) and it therefore remained at the originally assigned LoE1a.

We provide further examples in the supporting information, together with the detailed quality checklist filled in for each study (Table S1, S2 and S3).

Relevance for different user groups

In the previous section we have elaborated *how* to employ the evidence-based concept. Now we want to provide a few notes on *who* should use it:

1. Scientists conducting their own studies have to be aware how to achieve the highest possible evidence, particularly during the planning phase. Choosing a study design that provides a good evidence and respects quality criteria will substantially increase the potential contribution to our knowledge.

2. Scientists advising decision-makers should be aware of the evidence of information they include in their recommendations. Weighting all scientific information equally, or subjectively, runs the risk of overconfidence and bias.

3. Decision-makers receiving information from scientists should demand a level-of-evidence statement for the information provided, or should judge themselves the reliability having in mind the evidence-based concept.

4. We further would like to encourage **consortia, international panels and learned societies, such as the Intergovernmental Platform on Biodiversity & Ecosystem Services (IPBES) and Ecological Societies (ESA, BES, INTECOL)** to develop guidelines with recommendations on methods to best quantify, value, manage or govern a desired ecosystem service. This would give decision-makers more transparent advise, making the common procedure ('decision-makers seeking advice from individual scientists') a phase-out model. These 'best practice guides' ideally exist for every single and for the sum of ecosystem services in every focus and in every ecosystem. For example we may want to ask what is the best way to quantify recreation, to value recreation, to manage recreation and to use governance strategies that fosters sustainable recreation in a temperate forest. Each best practice guide would clearly state its level of evidence. All this may sound unrealistic, given the huge number of methods, ecosystem services, management and governance options and so forth. However, in medicine, national and international learned societies have set up assessment and guideline boards for exactly this purpose (often with governmental support, e.g. the US Agency for Healthcare Research and Quality (AHRQ) www.ahrq.gov, UK's National Institute for Health and Care Excellence (NICE) www.nice.org.uk or Germany's (IQWiG) www.iqwig.de). There are currently

261 recognised diseases with over 12000 sub-categories (listed in the ICD-10). This is certainly at the same scale as the challenges faced by ecosystem service science.

Conclusion

We outlined the evidence-based concept to ecosystem service science, encompassing a scale to judge the available evidence and a quality checklist to facilitate critical appraisal. We further showed in detail and illustrated with examples how to use the concept (see also supporting information). Additional support and guidance can be obtained by the Collaboration of Environmental Evidence (www.environmentalevidence.org).

The evidence-based concept does not suggest a specific management strategy. It is by no means a contradiction to or replacement of adaptive management or other management concepts. Rather, it complements these approaches, emphasising that whatever is used should be used with the awareness of how reliable our knowledge is.

Wrong decisions can have strong negative consequences. This is particularly painful, if studies providing high evidence were available but decisions were based instead on expert guesses or low-evidence studies. Child mortality from sudden infant death syndrome was unnecessary high for decades due to wrong recommendations based on low evidence, ignoring the higher evidence available at that time (Gilbert *et al.*, 2005). Especially on topics with contradicting opinions, it is important to continuously summarise and update the available evidence. If farmers have no reliable information on the management of natural pest control versus pesticides (Wright *et al.*, 2013), their actions may result in huge and avoidable economic loss or even directly affect human health.

It should have become clear that evidence-based ecosystem service science concerns scientists as well as decision-makers and the general public. In the interest of a responsible use of environmental resources and processes, we strongly encourage embracing evidence-based practice as paradigm for all research contributing to ecosystem service science.

Acknowledgements

We thank Sven Lautenbach, Ian Bateman, Jessica Walsh and Bill Sutherland for valuable comments on earlier versions of the manuscript. This work was supported by the 7th framework programme of the European Commission in the project OPERAs (grant number 308393, www.operas-project.eu).

References

Acuña V, Díez JR, Flores L, *et al.* 2013. Does it make economic sense to restore rivers for their ecosystem services? *Journal of Applied Ecology* **50**: 988–997.

Bilotta GS, Milner AM, and Boyd IL. 2014. Quality assessment tools for evidence from environmental science. *Environmental Evidence* **3**: 14.

Binkley D and Menyailo O. 2005. Gaining insights on the effects of tree species on soils. In: Tree Species Effects on Soils: Implications for Global Change, chapter 1, 1–16. Dordrecht: Kluwer Academic Publishers.

Bowler D, Buyung-Ali L, Healey JR, *et al.* 2010. The evidence base fo community forest management as a mechanism for supplying global environmental benefits and improving local welfare. *Collaboration for Environmental Evidence* **011**.

Bowler DE, Buyung-ali LM, Healey JR, *et al.* 2012. Does community forest management provide global environmental benefits and improve local welfare? *Frontiers in Ecology and the Environment* **10**.

Boyd I. 2013. A standard for policy-relevant science. *Nature* **501**: 159–160.

Collaboration for Environmental Evidence. 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Technical Report March, Environmental Evidence.

Daily GC, Polasky S, Goldstein J, *et al.* 2009. Ecosystem services in decision making: time to deliver. *Frontiers in Ecology and the Environment* **7**: 21–28.

Dicks LV, Showler DA, and Sutherland WJ. 2010. Bee conservation: evidence for the effects of interventions.

Dicks LV, Walsh JC, and Sutherland WJ. 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. *Trends in ecology & evolution* 1–7.

Entenmann SK and Schmitt CB. 2013. Actors' perceptions of forest biodiversity values and policy issues related to REDD plus implementation in Peru. *Biodiversity and Conservation* **22**: 1229–1254.

Gilbert R, Salanti G, Harden M, and See S. 2005. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology* **34**: 874–87.

GRADE Working Group. 2004. Grading quality of evidence and strength of recommendations. *BMJ* **328**: 1–8.

Graham R, Mancher M, Wolmann DM, *et al.* 2011. Clinical Practice Guidelines We Can Trust. Washington, DC: The National Academies Press.

Higgins JPT and Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions.

Higgs J and Jones M. 2000. Will evidence-based practice take the reasoning out of practice? In: Higgs J and Jones M (Eds.) Clinical Reasoning in the Health Professionals, 307–315. Oxford: Butterworth Heineman, 2 edition.

ICD-10. 2010. International Statistical Classification of Diseases and Related Health Problems, volume 2. World Health Organization.

Lawton JH. 1999. Are There General Laws in Ecology? *Oikos* **84**: 177–192.

Millennium Ecosystem Assessment. 2005. Ecosystems and Human Well- being: Synthesis. Technical report, published, Washington (DC): Island Press.

Nelson E, Mendoza G, Regetz J, *et al.* 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment* **7**: 4–11.

OCEBM Levels of Evidence Working Group. 2011. The Oxford Levels of Evidence 1. URL http://www.cebm.net/index.aspx?o=5653.

Petrokofsky G, Holmgren P, and Brown ND. 2011. Reliable forest carbon monitoring-systematic reviews as a tool for validating the knowledge base. *International Forestry Review* **13**: 56–66.

Pullin AS and Knight TM. 2001. Effectiveness in conservation practice: Pointers from medicine and public health. *Conservation Biology* **15**: 50–54.

Pullin AS and Knight TM. 2003. Support for decision making in conservation practice: an evidence-based approach. *Journal for Nature Conservation* **11**: 83–90.

Pullin AS and Knight TM. 2009. Doing more good than harm - Building an evidence-base for conservation and environmental management. *Biological Conservation* **142**: 931–934.

Pullin AS, Knight TM, and Watkinson AR. 2009. Linking reductionist science and holistic policy using systematic reviews: unpacking environmental policy questions to construct an evidence-based framework. *Journal of Applied Ecology* **46**: 970–975.

Raffaelli D and White PCL. 2013. Ecosystems and Their Services in a Changing World. Elsevier.

Sackett DL, Rosenberg WMC, Gray JAM, *et al.* 1996. Evidence based medicine: what it is and what it isn't. *Clinical Orthopaedics and Related Research* **455**: 3–5.

Söderqvist T and Soutukorva A. 2006. An instrument for assessing the quality of environmental valuation studies. URL http://www.naturvardsverket.se/Documents/publikationer/620-1252-5.pdf.

Sutherland WJ. 2000. The Conservation Handbook: Research, Management and Policy. Wiley.

Sutherland WJ, Pullin AS, Dolman PM, and Knight TM. 2004. The need for evidence-based conservation. *Trends In Ecology & Evolution* **19**: 305–8.

Tallis H and Polasky S. 2009. Mapping and Valuing Ecosystem Services as an Approach for Conservation and Natural-Resource Management. *Annals of the New York Academy of Sciences* **1162**: 265–283.

Tetlock PE. 2005. Expert Political Judgment: How Good Is It? How Can We Know? Princeton University Press.

UK Civilservice. 2013. What is a Rapid Evidence Assessment? URL http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is.

Vetter D, Rücker G, and Storch I. 2013. Meta-analysis: A need for well-defined usage in ecology and conservation biology. *Ecosphere* **4**: art74.

Wright H, Ashpole JE, Dicks LV, *et al.* 2013. Enhancing natural pest control as an ecosystem service: Evidence for the effects of selected actions. Cambridge: University of Cambridge.

Table 1. Disciplinary focus of an ecosystem service question

Focus	Description
Quantification	Amount of an ecosystem service or a set of services. It can be measured in absolute units or relative to another system.
Valuation	Societal value of a service or a set of services. The most common way is monetary valuation. Other possibilities are in relation to a reference system or on a ranked scale (high, middle, low value).
Management	Management/treatment of an ecosystem to favour specific ecosystem services. For example: leaving dead wood in forests to increase biodiversity or reducing agricultural fertiliser to decrease nearby lake eutrophication.
Governance	Strategy to steer a management type. The tools used by policy makers include e.g. incentives (subsidiaries) and/or penalties (law/tax).

Table 2. This **quality checklist** is designed in the form of questions. Each question answered with 'yes' will receive a point, important aspects (bold type) two points. If a question is not appropriate, it may be left out.

General aspects	
1	Does the question match the answer?
2	Are the assumptions used in the study reasonable?
3	Internal validity: Do design and implementation avoid a high risk of bias?
4	External validity/relevance: Is the result transferable to similar scenarios?
Data collection	
5	Was the target population/area defined in space, time and size?
6	Was a sampling population/area defined?
7	Were potential differences between the target population and the sampling population considered?
8	Were the methods described in sufficient detail to permit replication?
9	Was the sample size appropriate?
10	Was probability sampling used for constructing the sample?
11	If secondary data are used, did an evaluation of the original data take place?
Analysis	
12	Is the choice of statistical/analytical methods justified and comprehensively explained?
13	Are variables and statistical measures given?
14	Was accuracy/uncertainty assessed and reported?
Results	
15	Are results consistent and homogeneous?
16	Magnitude of effect: Is the effect large (and without large uncertainty)?
17	Is the outcome report complete and no information is missing?
18	Attrition bias: Are non-response/drop-outs given and is their impact discussed?
Design-specific aspects:	
Review	
19	Is there a low probability of publication bias? E.g. results reporting a negative relationship were probably not included
20	Is the review based on high evidence individual studies?
21	Do the studies included respond to the same question?
22	Was the literature searched in a systematic way?
23	Was a meta-analysis (in the strict sense: see Vetter et al. 2013) included?
24	Were appropriate study inclusion/exclusion criteria defined?
Study with a reference	
25	Selection bias: Was the assignment of case-control groups randomized?

26 Were groups designed equally, aside from the investigated point of interest?

27 Performance bias: Was the sampling blinded, e.g. researchers taking samples of a specific area wouldn't know the differences between these areas?

28 Were there sufficient replicates of treatment and reference units?

29 Detection bias: Were outcomes measured identically between groups?

Focus-specific aspects:

Quantification

30 Is the unit of the quantification measurement appropriate?

31 Was temporal change of ecosystem services' quantities (e.g. annual or long-term) discussed?

Valuation

32 Were future values of ecosystem services considered?

33 If future values were considered, were they discounted with a well-motivated discount rate?

34 If aggregate economic values for a population were estimated, was this estimation consistent with the sampling and the definition of the population?

35 If valuation took place in form of a questionnaire, was the study pre-tested and piloted?

Management

36 Was the aim of the management intervention clearly defined?

37 Were both long-term and short-term effects discussed?

38 Did monitoring take place for an appropriate time period?

39 Were stakeholders included?

40 Was the role of stakeholders described in detail?

Governance

41 Were long-term effects assessed?

42 Was the policy instrument that was used described and well chosen?

43 Was the influence of the policy instrument (incentive/law) on society discussed?

Table 3. Downgrading according to the percentage of quality points reached

Percentage of points	Level of shortcomings	Downgrading of the LoE
$>$ 75% of total points	–> no shortcomings	–> no downgrading
50 - 75% of total points	–> shortcomings	–> downgrading by one level
25 - 50% of total points	–> serious shortcomings	–> downgrading by two levels
$<$ 25% of total points	–> very serious shortcomings	–> downgrading by three levels

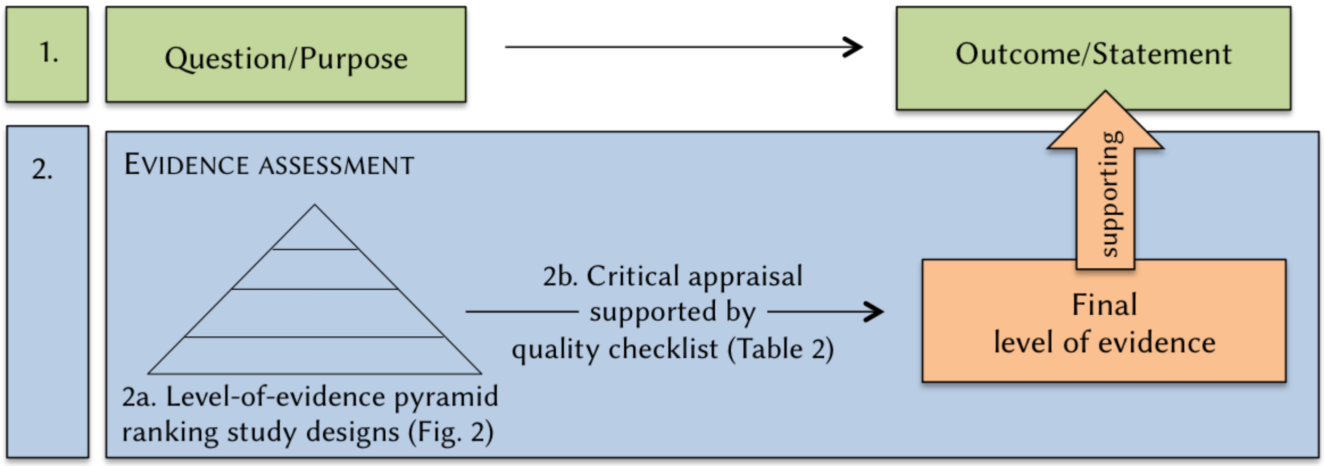

Figure 1. Schematic procedure of evidence-based practice: 1. Identification of the study question and the outcome, given as result of the study. 2. The assessment of the evidence supporting the outcome, with help of a level-of-evidence pyramid and a quality checklist.

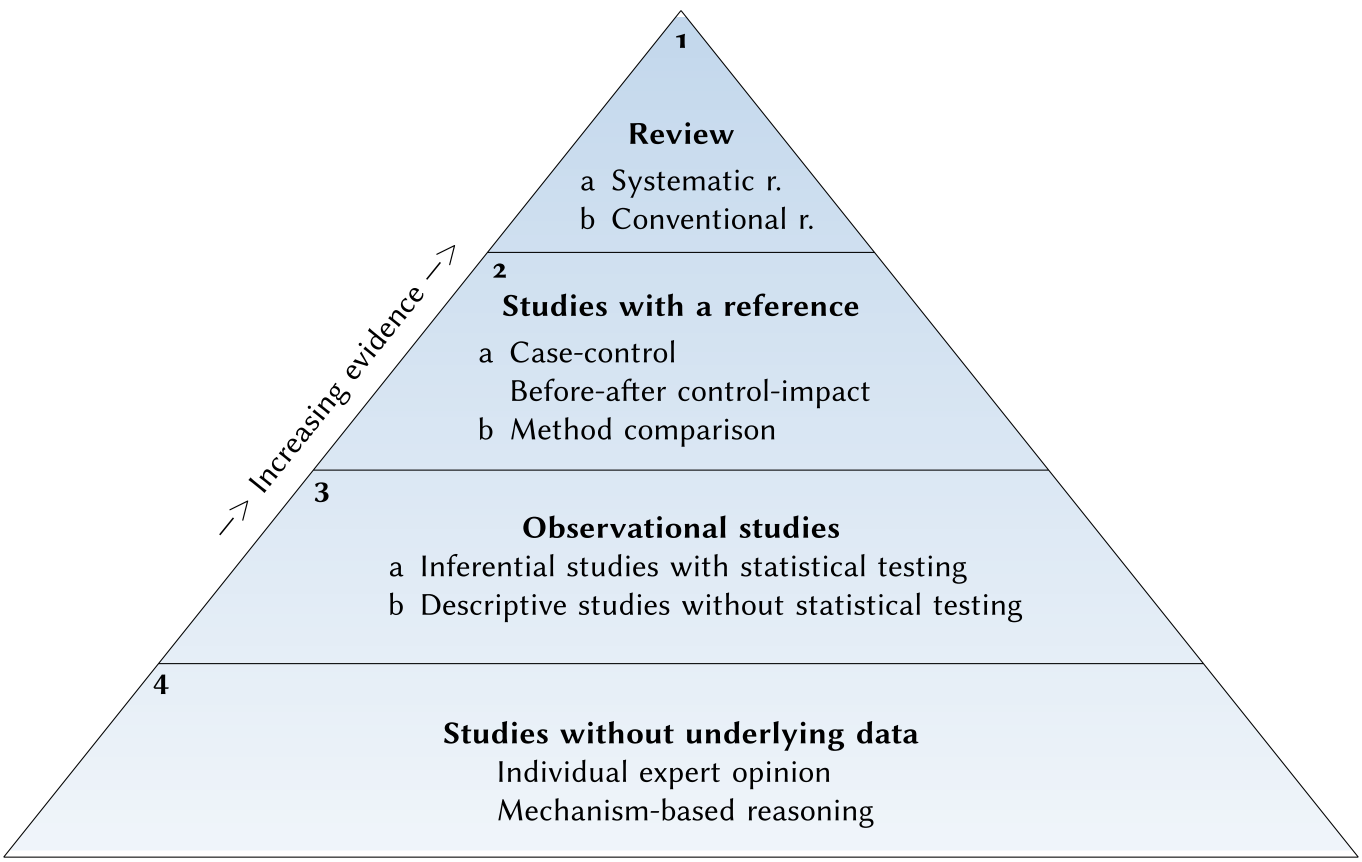

Figure 2. Level-of-evidence (LoE) pyramid ranking study designs according to their evidence. LoE1 - LoE4 with internally ranked subcategories a and b.