

An evidence assessment tool for ecosystem services and conservation studies

Supporting information

Anne-Christine Mupepele¹² & Jessica C. Walsh³
& William J. Sutherland³ & Carsten F. Dormann¹

Quality checklist

The quality checklist was composed out of 30 quality checklists, most of them with a focus on medical issues. Some quality criteria were added without being mentioned in previous checklists. This was especially the case for environmental management aspects that have not been addressed in checklists before. The weighting of the quality criteria was performed by the authors separately and brought to a consensus.

¹Department of Biometry and Environmental System Analysis, University of Freiburg, Tennenbacherstr. 4, 79106 Freiburg, Germany

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

³Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK.

WebTable 1: Quality checklist with references

Quality checklist question	Source of quality checklist criterion
INTERNAL VALIDITY	
<i>Research aim</i>	
1 Does the study address a clearly focused question?	Spencer <i>et al.</i> 2003, Lohr 2004, SIGN 2006, CEBM 2010, Collaboration for Environmental Evidence 2013
2 Does the question match the answer?	
<i>Data collection</i>	
3 Was the population/area of interest defined in space, time and size?	Spencer <i>et al.</i> 2003, Lohr 2004, Söderqvist & Soutukorva 2006, Brouwers <i>et al.</i> 2010, Santaguida <i>et al.</i> 2012, AHRQ 2014
4 Selection bias: Was the sample area representative for the population defined?	National Health and Medical Research Council 2000, Söderqvist & Soutukorva 2006, Tong <i>et al.</i> 2007, Moher <i>et al.</i> 2010, Santaguida <i>et al.</i> 2012
5 Was the sample size appropriate?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, Verhagen <i>et al.</i> 1998, Söderqvist & Soutukorva 2006, Tong <i>et al.</i> 2007, Moher <i>et al.</i> 2012, AHRQ 2014
6 Was probability/random sampling used for constructing the sample?	Söderqvist & Soutukorva 2006
7 If secondary data were used, did an evaluation of the original data take place?	Söderqvist & Soutukorva 2006
8 If data collection took place in form of a questionnaire, was it pre-tested/piloted?	Söderqvist & Soutukorva 2006, Rattray & Jones 2007, Tong <i>et al.</i> 2007
9 Were the data collection methods described in sufficient detail to permit replication?	Brouwers <i>et al.</i> 2010, CEBM 2010, Moher <i>et al.</i> 2010
<i>Analysis</i>	
10 Were the statistical/analytical methods described in sufficient detail to permit replication?	Lohr 2004, Brouwers <i>et al.</i> 2010, CEBM 2010, Moher <i>et al.</i> 2010
11 Is the choice of statistical/analytical methods appropriate and/or justified?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, Söderqvist & Soutukorva 2006
12 Was uncertainty assessed and reported?	Ah-See & Molony 1998, Söderqvist & Soutukorva 2006, Bastuji-Garin <i>et al.</i> 2013
<i>Results and Conclusions</i>	
13 Do the data support the outcome?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998
14 Magnitude of effect: Is the effect large, significant and/or without large uncertainty?	Jadad <i>et al.</i> 1996, Rychetnik <i>et al.</i> 2001, SIGN 2006, CEBM 2010, Singh <i>et al.</i> 2012
15 Are all variables and statistical measures reported?	CEBM 2010, Higgins <i>et al.</i> 2011, Bilotta <i>et al.</i> 2014
16 Attrition bias: Are non-response/drop-outs given and is their impact discussed?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, SIGN 2006, Söderqvist & Soutukorva 2006, Tong <i>et al.</i> 2007, Bilotta <i>et al.</i> 2014
DESIGN-SPECIFIC ASPECTS	
<i>Review</i>	
17 Is there a low probability of publication bias?	National Health and Medical Research Council 2000, SIGN 2006, Shea <i>et al.</i> 2007, CEBM 2010, AHRQ 2014
18 Is the review based on several strong-evidence individual studies?	SIGN 2006
19 Do the studies included respond to the same question?	AHRQ 2014
20 Are results between individual studies consistent and homogeneous?	Rychetnik <i>et al.</i> 2001, SIGN 2006, CEBM 2010
21 Was the literature searched in a systematic and comprehensive way?	SIGN 2006, Shea <i>et al.</i> 2007, Brouwers <i>et al.</i> 2010
22 Was a meta-analysis included?	
23 Were appropriate a priori study inclusion/exclusion criteria defined?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, Verhagen <i>et al.</i> 1998, Lohr 2004, Shea <i>et al.</i> 2007, CEBM 2010, Tong <i>et al.</i> 2012, Moher <i>et al.</i> 2014
24 Did at least two people select studies and extract data?	SIGN 2006, Shea <i>et al.</i> 2007, CEBM 2010
<i>Study with a reference/control</i>	
25 Allocation bias: Was the assignment of case-control groups randomized?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, Verhagen <i>et al.</i> 1998, National Health and Medical Research Council 2000, Lohr 2004, SIGN 2006, CEBM 2010, Moher <i>et al.</i> 2010, Higgins <i>et al.</i> 2011
26 Were groups designed equally, aside from the investigated point of interest?	Lohr 2004, SIGN 2006, CEBM 2010
27 Performance bias: Was the sampling blinded?	Jadad <i>et al.</i> 1996, Ah-See & Molony 1998, Verhagen <i>et al.</i> 1998, Rychetnik <i>et al.</i> 2001, Lohr 2004, SIGN 2006, CEBM 2010, Moher <i>et al.</i> 2010, Higgins <i>et al.</i> 2011, Bilotta <i>et al.</i> 2014
28 Were there sufficient replicates of treatment and reference groups?	SIGN 2006
29 Detection bias: Were outcomes equally measured and determined between groups?	Bilotta <i>et al.</i> 2014
Observational studies	
30 Were confounding factors identified and strategies to deal with them stated?	Joanna Briggs Institute 2014

WebTable 1: Quality checklist with references

Quality checklist question	Source of quality checklist criterion
FOCUS-SPECIFIC ASPECTS	
<i>Quantification</i>	
31	Is the unit of the quantification measurement appropriate?
32	Was temporal change (e.g. annual or long-term) of quantities measured (e.g. species abundance or an ecosystem service) discussed?
<i>Valuation</i>	
33	If discounting of future costs and outcomes is necessary, was it performed correctly?
	SIGN 2006, Söderqvist & Soutukorva 2006
34	If aggregate economic values for a population were estimated, was this estimation consistent with the sampling and the definition of the population?
	Defra 2007, de Groot <i>et al.</i> 2012
<i>Management</i>	
35	Was the aim of the management intervention clearly defined?
36	Were side effects and trade offs on other non-target species, ecosystem services or stakeholders considered?
37	Were both long-term and short-term effects discussed?
	AHRQ 2014
38	Did monitoring take place for an appropriate time period?
	Jadad <i>et al.</i> 1996, CEBM 2010
39	Appropriate outcome measures: Are all relevant outcomes measured in a reliable way?
	Jadad <i>et al.</i> 1996, SIGN 2006
<i>Governance</i>	
40	Were long-term effects assessed?
	Biermann & Pattberg 2012, ARHQ 2014
41	Was the policy instrument that was used described?
42	Was the influence of the applied policy instrument (incentive/law) on the society discussed?
43	Appropriate outcome measures: Are all relevant outcomes measured in a reliable way?
	Jadad <i>et al.</i> 1996, SIGN 2006

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Mant <i>et al.</i> 2013	Lindhjem 2007	Liu <i>et al.</i> 2008	Acuna <i>et al.</i> 2013
Context: Subject/Ecosystem services; Ecosystem(s); Location	Fish and aquatic invertebrates; freshwater; global	Non-timber benefits (mainly recreation); boreal forests; Norway, Sweden, Finland	Timber, soil erosion, carbon sequestration, recreation; forests; China	Fish, recreation, erosion control; stream; Spain
Focus (Quantification, Valuation, Management or Governance)	Management	Valuation	Governance	Management
Question/Purpose investigated	What is the impact of 'liming' (adding Calcium carbonate) of streams and rivers on the abundance and diversity of fish and invertebrate populations?	How to explain systematic variation in Willingness-to-Pay (WTP) for the value of non-timber benefits from forests in Fennoscandia?	What is the socioeconomic and ecological impact of two payments-for-ecosystem-services programs in China?	Does adding dead wood to streams affect the value of selected ecosystem services and is it cost-effective?
Outcome	Liming increased fish abundances and acid sensitive invertebrates, but effects were variable and for all invertebrate taxa combined liming may decrease abundance.	WTP is insensitive to the size of the forest and tends to be higher if individuals are asked instead of households.	Socioeconomic impact: income increased, but revenues declined for local governments. Ecological impact: Timber harvest decreased locally but import increased. Carbon sequestration increased and soil erosion declined.	Restoration of natural wood loading in streams increases the ecosystem service provision. The cost-benefit analysis reveals differences between stream orders in the net benefit of the restoration.
2a. Study design	Systematic Review	Review	Review	BACI
Level of evidence	LoE1a	LoE1b	LoE1b	LoE2a
2 b. QUALITY CHECKLIST FOR THE CRITICAL APPRAISAL				
INTERNAL VALIDITY	Description/Example	Quality checklist	Quality checklist	Quality checklist
Research aim		Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"
1 Does the study address a clearly focused question?	See main text, section 'setting question and the context' Answers may not directly correspond to the originally formulated question, e.g. 'Does hunting lead to genetic changes in the moose population of North America?' is answered by: 'hunting reduces the size of calves'. The missing match is obvious when question and answer are written next to each other, but in publications with much text in between it may be more difficult to identify. The result of reduced calf size may be interesting, but special care should be taken while assessing the evidence base.	yes	yes	no
2 Does the question match the answer?	Usually samples are not taken from the whole population/area; e.g. only several North American forests were selected to measure moose. Were the selected forests representative? Did they cover the north, south, east and western part of North America?	yes	yes	yes
Data collection				
3 Was the population/area of interest defined in space, time and size?	'Population/area' is the target, we aim to say something about; e.g. North America's moose population.	no	yes	yes
4 Selection bias: Was the sample area representative for the population defined?	Were the criteria used to determine the sample size (e.g. power calculation) reasonable?	/	yes	yes
5 Was the sample size appropriate?	Probability sampling means random sampling with known selection probabilities for all objects in the population, while nonprobability sampling does not involve random selection (Trochim, 2014; Söderqvist and Soutukorva, 2009). Most often <i>equal</i> probability sampling is used: e.g. all forests in North America have the same chance of being randomly selected. <i>Unequal</i> probability sampling can be used to ensure representativeness of result, e.g. if a forest in the south of the area is selected, the selection of the next forest far away from the first will be favored. Unequal probability sampling can also mean that forests easy to access obtain a higher selection probability. Probability sampling is important in addition to representative sampling (question 4).	yes	yes	/
6 Was probability/random sampling used for constructing the sample?	Secondary data, such as used in cost-benefit transfer for example, need to be evaluated to make sure that the data used are not prone to bias. Questionnaires need to be professionally designed to ensure that they measure what they intend to measure. Therefore a questionnaire should be pre-tested/piloted on a smaller sample size to test its performance (see Rattray & Jones 2007).	/	/	/
7 If secondary data were used, did an evaluation of the original data take place?	Were the data collection methods described in sufficient detail to permit replication?	yes	no	no
8 If data collection took place in form of a questionnaire, was it pre-tested/piloted?	Were the statistical/analytical methods described in sufficient detail to permit replication?	/	/	/
9 Were the data collection methods described in sufficient detail to permit replication?	Is the choice of statistical/analytical methods appropriate and/or justified?	yes	no	yes
Analysis				
10 Were the statistical/analytical methods described in sufficient detail to permit replication?	Was uncertainty assessed and reported?	yes	yes	no
11 Is the choice of statistical/analytical methods appropriate and/or justified?	Results and Conclusions	yes	yes	yes
12 Was uncertainty assessed and reported?	Do the data support the outcome?	yes	yes	/
13 Do the data support the outcome?	Are the conclusions drawn of the analytical results valid?	yes	yes	yes
14 Magnitude of effect: Is the effect large, significant and/or without large uncertainty?	This question aims to identify the magnitude and precision of results. Precise results are usually characterized by low uncertainty (CEBM 2010) and in combination with a large effect the appropriate statistical analysis (question 11) will lead to a significant result. Not all studies allow the judgment of all three aspect and we therefore combine them in one question and recommend context specific decisions.	yes	yes	/
15 Are all variables and statistical measures reported?	Attrition bias: Are non-response/drop-outs given and is their impact discussed?	yes	yes	/
16 Attrition bias: Are non-response/drop-outs given and is their impact discussed?		/	yes	/

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Mant <i>et al.</i> 2013	Lindhjem 2007	Liu <i>et al.</i> 2008	Acuna <i>et al.</i> 2013
DESIGN-SPECIFIC ASPECTS				
Review				
17	no	yes	no	/
18	yes	yes	no	/
19	yes	yes	/	/
20	yes	no	yes	/
21	yes	yes	no	/
22	yes	yes	no	/
23	yes	no	no	/
24	yes	no	no	/
25	/	/	/	no
26	/	/	/	yes
27	/	/	/	no
28	/	/	/	yes
29	/	/	/	yes
30	/	/	/	/
FOCUS-SPECIFIC ASPECTS				
Quantification				
31	/	/	/	/
32	/	/	/	/
33	/	no	/	no
34	/	/	/	/
Management				
35	yes	/	/	yes
36	no	/	/	no
37	yes	/	/	no
38	/	/	/	yes
39	yes	/	/	yes
Governance				
40	/	/	yes	/
41	/	/	yes	/
42	/	/	yes	/
43	/	/	yes	/
2b. Quality points	25	22	10	22
Possible points (depending on the number of questions answered)	28	28	21	31
Quality score	89.29	78.57	47.62	70.97
Downgrading	no downgrading	half a level	two levels	one level
Level of evidence	LoE 1a	LoE 2a	LoE 3b	LoE 3a

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Kleijn <i>et al.</i> 2006	Millar <i>et al.</i> 2010	Bastin <i>et al.</i> 2014	Goulson <i>et al.</i> 2002	Rundio and Olson 2007	
1. Question, outcome and the context	Context: Subject/Ecosystem	Biodiversity (vascular plants, birds, bees, grasshoppers, crickets, spiders); farmland; Europe	Soil; grassland; USA	Biomass; tropical forests; Congo	<i>Bombus terrestris</i> ; farmland, suburban area; UK	Salamanders; forests; USA
	Focus (Quantification, Valuation)	Governance	Quantification	Quantification	Management	Management
	Question/Purpose investigated	Do agri-environment schemes have an effect on biodiversity and endangered species?	Does commercial soil production result in net soil loss? Is there a way to measure the natural occurring soil that is lost with each harvest?	Demonstrating the feasibility to create an aboveground biomass map through a regional study of canopy texture by harmonizing Fourier textural ordination (FOTO) indices of images.	Do measures to promote farmland biodiversity have an influence on nest growth of <i>Bombus terrestris</i> ?	What are the short-term effects of forest thinning on terrestrial salamanders in managed headwater forests? Can down wood or riparian buffers influence these effects?
	Outcome	Agri-environmental schemes had marginal to moderately positive effects on biodiversity, but endangered species rarely benefit.	Yes. There is a net soil loss of around 100 Mg per year, which is considerably higher than the tolerable soil loss.	Good agreement was found between observed and predicted aboveground biomass and a high-resolution biomass map was produced for a 400km ² area in the Congo basin.	Schemes deployed to enhance farmland biodiversity appear to have little measurable impact on nest growth of this bumblebee species.	Forest thinning decreases salamander abundance in forests that have a low down-wood volume. In stands with little down wood, riparian buffer width would need consideration and may help minimize negative effects of thinning on salamanders.
2a. Study design	Case control	Case control	Method comparison	Case control	BACI	
Level of evidence	LoE2a	LoE2a	LoE2b	LoE2a	LoE2a	
2 b. QUALITY CHECKLIST	Quality checklist	Quality checklist	Quality checklist	Quality checklist	Quality checklist	
INTERNAL VALIDITY	Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"	
Research aim						
1 Does the study address a clearly focused question?	yes	yes	no	no	yes	
2 Does the question match the answer?	yes	yes	yes	yes	yes	
Data collection						
3 Was the population/area of interest defined in space, time and size?	yes	yes	yes	yes	yes	
4 Selection bias: Was the sample area representative for the population defined?	yes	no	yes	no	yes	
5 Was the sample size appropriate?	yes	yes	yes	yes	no	
6 Was probability/random sampling used for constructing the sample?	no	no	yes	yes	no	
7 If secondary data were used, did an evaluation of the original data take place?	/	/	no	/	/	
8 If data collection took place in form of a questionnaire, was it pre-tested/piloted?	/	/	/	no	/	
9 Were the data collection methods described in sufficient detail to permit replication?	yes	yes	yes	yes	no	
Analysis						
10 Were the statistical/analytical methods described in sufficient detail to permit replication?	yes	yes	yes	yes	no	
11 Is the choice of statistical/analytical methods appropriate and/or justified?	yes	yes	yes	yes	no	
12 Was uncertainty assessed and reported?	yes	no	no	yes	no	
Results and Conclusions						
13 Do the data support the outcome?	yes	yes	yes	no	yes	
14 Magnitude of effect: Is the effect large, significant and/or without large uncertainty?	no	yes	/	no	no	
15 Are all variables and statistical measures reported?	yes	yes	yes	yes	no	
16 Attrition bias: Are non-response/drop-outs given and is their impact discussed?	yes	/	/	yes	/	

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Kleijn <i>et al.</i> 2006	Millar <i>et al.</i> 2010	Bastin <i>et al.</i> 2014	Goulson <i>et al.</i> 2002	Rundio and Olson 2007
DESIGN-SPECIFIC ASPECTS					
Review					
17	/	/	/	/	/
18	/	/	/	/	/
19	/	/	/	/	/
20	/	/	/	/	/
21	/	/	/	/	/
22	/	/	/	/	/
23	/	/	/	/	/
24	/	/	/	/	/
Study with a reference/control					
25	no	no	/	no	no
26	no	yes	/	no	no
27	no	no	no	no	no
28	yes	yes	yes	yes	no
29	yes	yes	/	yes	yes
Observational studies					
30	/	/	/	/	/
FOCUS-SPECIFIC ASPECTS					
Quantification					
31	yes	yes	yes	yes	yes
32	no	yes	no	no	yes
Valuation					
33	/	/	/	/	/
34	/	/	/	/	/
Management					
35	/	/	/	no	yes
36	/	/	/	no	no
37	/	/	/	no	yes
38	/	/	/	no	yes
39	/	/	/	no	no
Governance					
40	no	/	/	/	/
41	yes	/	/	/	/
42	no	/	/	/	/
43	yes	/	/	/	/
2b. Quality points	21	19	16	16	13
Possible points (depending on Quality score)	31	26	21	33	31
Downgrading	67.74	73.08	76.19	48.48	41.94
Level of evidence	one level LoE3a	one level LoE3a	half a level LoE3a	two levels LoE4	two levels LoE4

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Entenmann and Schmitt 2013	Karimzadegan <i>et al.</i> 2007	Xie <i>et al.</i> 2011	Desanker 2005
Context: Subject/Ecosystem se	Biodiversity; forests; Peru	Gas regulation, pollination, pest control and other ecosystem services; forests; Iran	Air quality; urban area; China	Global climate regulation (C-sequestration); tropical forest; Africa
Focus (Quantification, Valuation)	Governance	Valuation	Quantification	Governance
Question/Purpose investigated	How do actors involved in REDD+ processes relate to biodiversity conservation? What aspects of biodiversity do they regard as especially important (biodiversity conservation values)?	What is the economic value of ecosystem services provided by Iran's forests and rangelands?	The air quality indicators: CO ₂ , O ₃ , SO ₂ , transpiration cooling and dust interception were quantified (and valued) for sixteen plant species.	How can the Clean Development Mechanism be better engaged in Africa?
Outcome	Biodiversity is not a major issue for actors, but direct synergies between REDD+ and biodiversity conservation were assumed by most actors. Values most often mentioned were direct or indirect use values. Option values for future benefits and resilience were rarely mentioned.	The economic value of nonmarket ecosystem services of forests and rangelands' is US\$ 53441 million annually. This is equivalent to 43% of Iran's GDP.	Plants with high leaf area indices and photosynthetic rates resulted in an increased transpiration cooling. Species with rough leaf surfaces are efficient in capturing dust and those with thick sclerophyllous leaves best remove SO ₂ .	Projects should be developed by locals. Carbon money alone may not be enough. Values from the services should be factored into the economic analysis of the country.
2a. Study design Level of evidence	Descriptive LoE3b	Descriptive LoE3b	Descriptive LoE3b	Expert opinion LoE4
2 b. QUALITY CHECKLIST	Quality checklist	Quality checklist	Quality checklist	Quality checklist
INTERNAL VALIDITY	Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"	Answer: "Yes/No"
Research aim				
1 Does the study address a clearly focused question?	yes	yes	yes	not required – already on lowest level of evidence
2 Does the question match the answer?	yes	yes	yes	
Data collection				
3 Was the population/area of interest defined in space, time and size?	yes	yes	yes	
4 Selection bias: Was the sample area representative for the population defined?	no	/	yes	
5 Was the sample size appropriate?	yes	/	yes	
6 Was probability/random sampling used for constructing the sample?	no	/	no	
7 If secondary data were used, did an evaluation of the original data take place?	/	no	/	
8 If data collection took place in form of a questionnaire, was it pre-tested/piloted?	no	/	/	
9 Were the data collection methods described in sufficient detail to permit replication?	yes	no	yes	
Analysis				
10 Were the statistical/analytical methods described in sufficient detail to permit replication?	yes	yes	yes	
11 Is the choice of statistical/analytical methods appropriate and/or justified?	yes	yes	yes	
12 Was uncertainty assessed and reported?	no	no	no	
Results and Conclusions				
13 Do the data support the outcome?	yes	yes	yes	
14 Magnitude of effect: Is the effect large, significant and/or without large uncertainty?	/	/	no	
15 Are all variables and statistical measures reported?	no	yes	yes	
16 Attrition bias: Are non-response/drop-outs given and is their impact discussed?	/	/	/	

WebTable 2: Evidence assessment tool applied to 13 case studies

Reference	Entenmann and Schmitt 2013	Karimzadegan <i>et al.</i> 2007	Xie <i>et al.</i> 2011	Desanker 2005
DESIGN-SPECIFIC ASPECTS				
Review				
17	/	/	/	
18	/	/	/	
19	/	/	/	
20	/	/	/	
21	/	/	/	
22	/	/	/	
23	/	/	/	
24	/	/	/	
Study with a reference/control				
25	/	/	/	
26	/	/	/	
27	/	/	/	
28	/	/	/	
29	/	/	/	
Observational studies				
30	no	no	no	
FOCUS-SPECIFIC ASPECTS				
Quantification				
31	/	/	yes	
32	/	/	no	
Valuation				
33	/	no	/	
34	/	no	/	
Management				
35	/	/	/	
36	/	/	/	
37	/	/	/	
38	/	/	/	
39	/	/	/	
Governance				
40	yes	/	/	
41	no	/	/	
42	yes	/	/	
43	yes	/	/	
2b. Quality points	13	8	14	
Possible points (depending on Quality score)	22	15	20	
Downgrading	59.09	53.33	70.00	
Level of evidence	one and a half levels LoE4	one and a half levels LoE4	one level LoE4	LoE4

WebTable 3: Studies on carbon sequestration (CS) in forests. Examples are given for each focus (quantification, valuation, management, governance) and all levels of evidence. No critical appraisal was performed, but this example highlights the use of the evidence hierarchy and the range of foci from quantification to governance. Carbon sequestration was a prominent topic over the previous years (Oren *et al.*, 2001; Fernández-Martínez *et al.*, 2014) and we found studies about carbon sequestration following different study designs. The studies vary in their geographical region and purpose of investigation. They may also investigate a broader range, e.g. the value of all ecosystem services, and we extracted only the question related to carbon sequestration.

	Quantification	Valuation	Management	Governance
Question:	How much carbon can be captured and stored by a forest?	What is the value of carbon sequestration in a forest?	How can we manage a forest to maximize carbon sequestration?	What are the best governance measures to manage a forest to maximize carbon sequestration?
Review (LoE1 if there are no quality shortcomings)	Does nutrient availability determine CS in forests? (Fernandez-Martinez et al. 2014)	What is the monetary value of CS provided by urban trees in Lisbon? (Roy, Byrne & Pickering 2012)	What is the effect of forest management on CS in soils? (Jandl et al. 2007)	How can we overcome critical challenges to scale up carbon investments in carbon sequestration projects in Africa? (Jindal, Swallow & Kerr 2008)
Referenced study (LoE2 if there are no quality shortcomings)	Does CS in forests depend on soil fertility? (Oren et al. 2001)	What is the non-market value from an afforested area in Spain? - Comparing results with contingent valuation and choice modelling (Mogas, Riera, Bennett 2006)	Impact of prescribed fire and small clear-cut tree harvesting on carbon dynamics in a mixed-conifer forest in Sierra Nevada? (Stephens et al. 2013)	What are barriers in implementing forest carbon trading? A comparison between the Clean Development Mechanism and a State-run carbon forestry program. (Corbera & Brown 2008)
Observational study (LoE3 if there are no quality shortcomings)	What is the reason for an increased CS in boreal deciduous forests in Canada between 1994 and 1998? (Black et al. 2000)	What is the value of CS provided by Canberra's urban forests? (Brack 2002)	Does carbon fixation increase with different forest management strategies (e.g. fertilization, thinning)? (Hoen 1994)	What are the effects of carbon taxes and subsidies on the supply of carbon services in West-Canada? (Van Kooten, Binkley & Delcourt 1995)
Based on no data (LoE4)	No study	No study	Does proper design and management of agroforestry result in effective carbon sinks? (Montagnini & Nair 2012)	What governance conditions have to be met to successfully put in practice small-scale forest carbon projects? (Boyd, Gutierrez & Chang 2007)

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.
- Ah-See KW and Molony NC. 1998. A qualitative assessment of randomized controlled trials in otolaryngology. *The Journal of Laryngology & Otology* **112**: 460–463.
- AHRQ - Agency for Healthcare Research and Quality. 2014. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville: MD: Agency for Healthcare Research and Quality.
- Bastin JF, Barbier N, Coueron P, *et al.* 2014. Aboveground biomass mapping of African forest mosaics using canopy texture analysis: toward a regional approach. *Ecological Applications* **24**: 1984–2001.
- Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, *et al.* 2013. Impact of STROBE Statement Publication on Quality of Observational Study Reporting: Interrupted Time Series versus Before-After Analysis. *PLoS ONE* **8**: 2–9.
- Bilotta GS, Milner AM, and Boyd IL. 2014. Quality assessment tools for evidence from environmental science. *Environmental Evidence* **3**: 14.
- Black TA, Chen WJ, Barr AG, *et al.* 2000. Increased carbon sequestration by a boreal deciduous forest in years with a warm spring. *Geophysical Research Letters* **27**: 1271–1274.
- Boyd E, Gutierrez M, and Chang M. 2007. Small-scale forest carbon projects: Adapting CDM to low-income communities. *Global Environmental Change* **17**: 250–259.
- Brack CL. 2002. Pollution mitigation and carbon sequestration by an urban forest. *Environmental pollution* **116**: 195–200.
- Brouwers M, Kho M, Browman GP, *et al.* 2010. AGREE II: Advancing guideline, development, reporting and evaluation in healthcare. *Canadian Medical Association Journal* **182**: 839–842.
- CEBM. 2010. Centre for Evidence-Based Medicine - Study Appraisal Worksheet. Oxford, UK.
- Collaboration for Environmental Evidence. 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2.
- Corbera E and Brown K. 2008. Building Institutions to Trade Ecosystem Services: Marketing Forest Carbon in Mexico. *World Development* **36**: 1956–1979.
- de Groot RS, Brander L, van der Ploeg S, *et al.* 2012. Global estimates of the value of ecosystems and their services in monetary units. *Ecosystem Services* **1**: 50–61.
- DEFRA - UK Department for Environment Food and Rural Affairs. 2007. An introductory guide to valuing ecosystem services. URL <https://www.gov.uk/government/publications/an-introductory-guide-to-valuing-ecosystem-services>.
- Desanker PV. 2005. The Kyoto Protocol and the CDM in Africa: a good idea but *Unasylva* **56**: 24–26.

- 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.
- Fernández-Martínez M, Vicca S, Janssens IA, *et al.* 2014. Nutrient availability as the key regulator of global forest carbon balance. *Nature Climate Change* **4**: 471–476.
- Goulson D, Hughes W, Derwent L, and Stout J. 2002. Colony growth of the bumblebee, *Bombus terrestris*, in improved and conventional agricultural and suburban habitats. *Oecologia* **130**: 267–273.
- Higgins JPT, Altman DG, Gøtzsche PC, *et al.* 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed)* **343**: d5928.
- Hoen HF. 1994. Potential and Economic Efficiency of Carbon Sequestration in Forest Biomass Through Silvicultural Management. *Forest Science* **40**: 429–451.
- Jadad AR, Moore RA, Carroll D, *et al.* 1996. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled clinical trials* **17**: 1–12.
- Jandl R, Lindner M, Vesterdal L, *et al.* 2007. How strongly can forest management influence soil carbon sequestration? *Geoderma* **137**: 253–268.
- Jindal R, Swallow B, and Kerr J. 2008. Forestry-based carbon sequestration projects in Africa: Potential benefits and challenges. *Natural Resources Forum* **32**: 116–130.
- Joanna Briggs Institute. 2014. Checklist for critical appraisal of descriptive studies. URL http://joannabriggs.org/assets/docs/jbc/operations/criticalAppraisalForms/JBC_Form_CritAp_DescCase.pdf.
- Karimzadegan H, Rahmatian H, Dehghani Salmasi M, *et al.* 2007. Valuing Forests and Rangelands-Ecosystem Services. *International Journal of Environmental Research* **1**: 368–377.
- Kleijn D, Baquero RA, Clough Y, *et al.* 2006. Mixed biodiversity benefits of agri-environment schemes in five European countries. *Ecology Letters* **9**: 243–54; discussion 254–7.
- Lindhjem H. 2007. 20 years of stated preference valuation of non-timber benefits from Fennoscandian forests: A meta-analysis. *Journal of Forest Economics* **12**: 251–277.
- Liu J, Li S, and Ouyang Z. 2008. Ecological and socioeconomic effects of China's policies for ecosystem services. *Proceedings of the National Academy of Sciences* **105**: 94779482.
- Lohr KN. 2004. Rating the strength of scientific evidence: relevance for quality improvement programs. *International Journal for Quality in Health Care* **16**: 9–18.
- Mant RC, Jones DL, Reynolds B, *et al.* 2013. A systematic review of the effectiveness of liming to mitigate impacts of river acidification on fish and macro-invertebrates. *Environmental Pollution* **179**: 285–293.
- Millar D, Stolt M, and Amador JA. 2010. Quantification and Implications of Soil Losses from Commercial Sod Production. *Soil Science Society of America Journal* **74**: 892–897.

- Mogas J, Riera P, and Bennett J. 2006. A comparison of contingent valuation and choice modelling with second-order interactions. *Journal of Forest Economics* **12**: 5–30.
- Moher D, Hopewell S, Schulz KF, *et al.* 2010. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical research ed)* **340**: 1–2.
- Moher D, Liberati A, Tetzlaff J, and Altman DG. 2014. Preferred Reporting Items for Systematic Reviews and Meta-Analyses. *Annals of Internal Medicine* **151**: 264–269.
- Montagnini F and Nair PKR. 2012. Carbon sequestration: An underexploited environmental benefit of agroforestry systems. *Agroforestry Systems* 281–295.
- National Health and Medical Research Council. 2000. How to review the evidence: systematic identification and review of the scientific literature. Government Info Bookshops.
- Oren R, Ellsworth DS, Johnsen KH, *et al.* 2001. Soil fertility limits carbon sequestration by forest ecosystems in a CO₂-enriched atmosphere. *Nature* **411**: 469–72.
- Rattray J and Jones MC. 2007. Essential elements of questionnaire design and development. *Journal of Clinical Nursing* **16**: 234–243.
- Roy S, Byrne J, and Pickering C. 2012. A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening* **11**: 351–363.
- Rundio DE and Olson DH. 2007. Influence of Headwater Site Conditions and Riparian Buffers on Terrestrial Salamander Response to Forest Thinning. *Forest Science* **53**: 320–330.
- Rychetnik L, Frommer M, Hawe P, and Shiell A. 2001. Criteria for evaluation evidence on public health interventions. *Journal of Epidemiology and Community Health* **56**: 119–127.
- Santaguida PL, Riley CM, and Matchar DB. 2012. Assessing risk of bias as a domain of quality in medical test studies. In: *Methods Guide for Medical Test Reviews*, chapter 5. Rockville: MD: Agency for Healthcare Research and Quality.
- Shea BJ, Grimshaw JM, Wells Ga, *et al.* 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology* **7**.
- SIGN. 2006. Reporting and quality checklists. URL www.sign.ac.uk.
- Singh S, Chang S, Matchar DB, and Bass EB. 2012. Grading a Body of Evidence on Diagnostic Tests. In: *Methods Guide for Medical Test Reviews*, chapter 7, 1–15. Rockville: MD: Agency for Healthcare Research and Quality.
- 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>.
- Spencer L, Ritchie J, Lewis J, and Dillon L. 2003. *Quality in Qualitative Evaluation: A framework for assessing research evidence*.

- Stephens S, Collins BM, Dore S, *et al.* 2013. Climate Change, Carbon Sequestration, and Wildfire Management in Sierran Mixed Conifer Forests. JFSP Research Project Reports.
- Tong A, Flemming K, McInnes E, *et al.* 2012. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC medical research methodology* **12**: 181.
- Tong A, Sainsbury P, and Craig J. 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care* **19**: 349–57.
- Trochim WMK. 2014. Research Methods Knowledge Base, 2nd edition. URL <http://www.socialresearchmethods.net/kb> (version current as of October 20, 2006).
- van Kooten GC, Binkley CS, and Delcourt G. 1995. Effect of Carbon Taxes and Subsidies on Optimal Forest Rotation Age and Supply of Carbon Services. *American Journal of Agricultural Economics* **77**: 365.
- Verhagen AP, de Vet HCW, de Bie RA, *et al.* 1998. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of clinical epidemiology* **51**: 1235–41.
- Xie QJ, Zhou ZX, and Chen F. 2011. Quantifying the Beneficial Effect of Different Plant Species On Air Quality Improvement. *Environmental Engineering and Management Journal* **10**: 959–963.