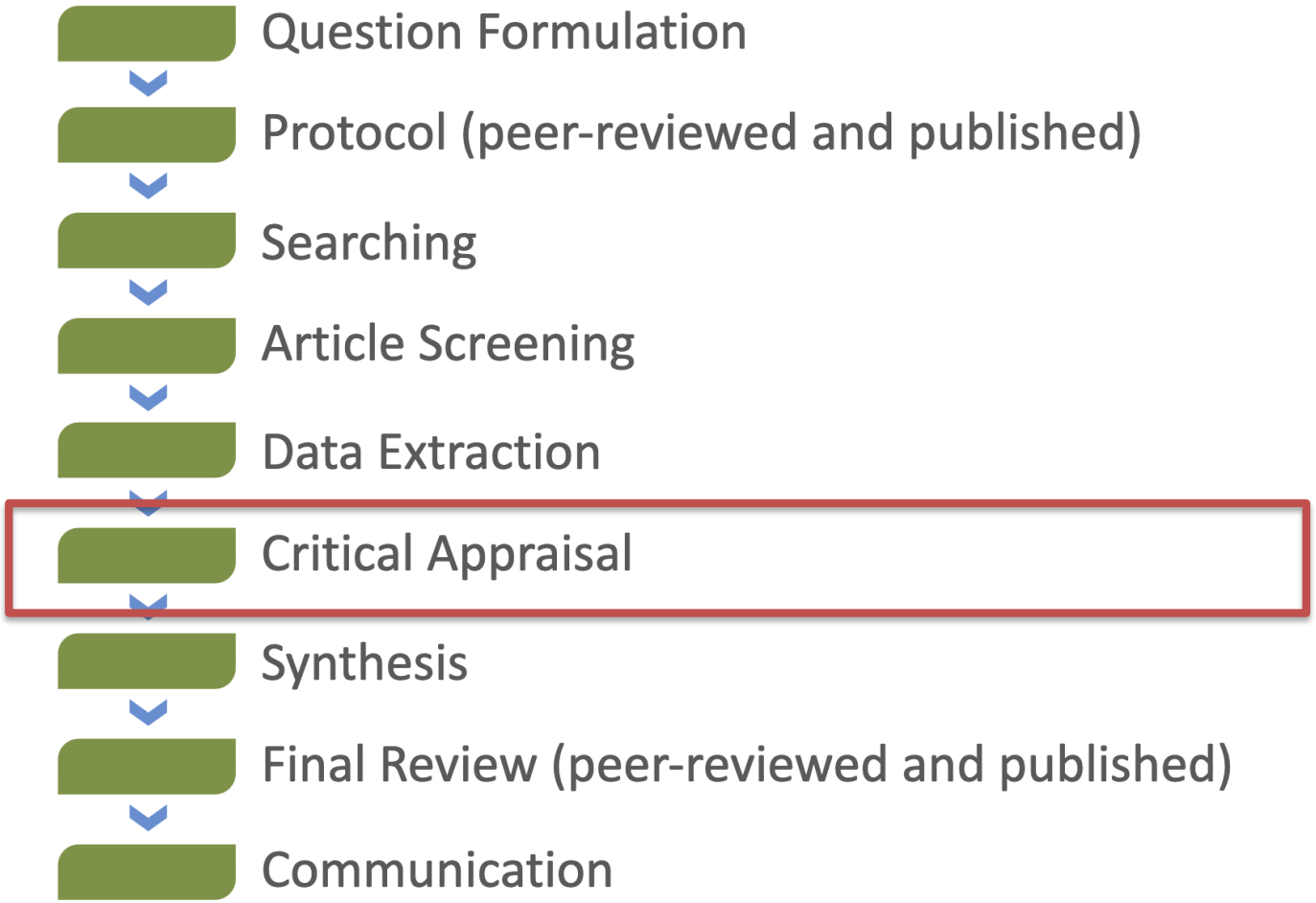
FRB
FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

CESAB
CENTRE DE SYNTHÈSE ET D'ANALYSE SUR LA BIODIVERSITÉ

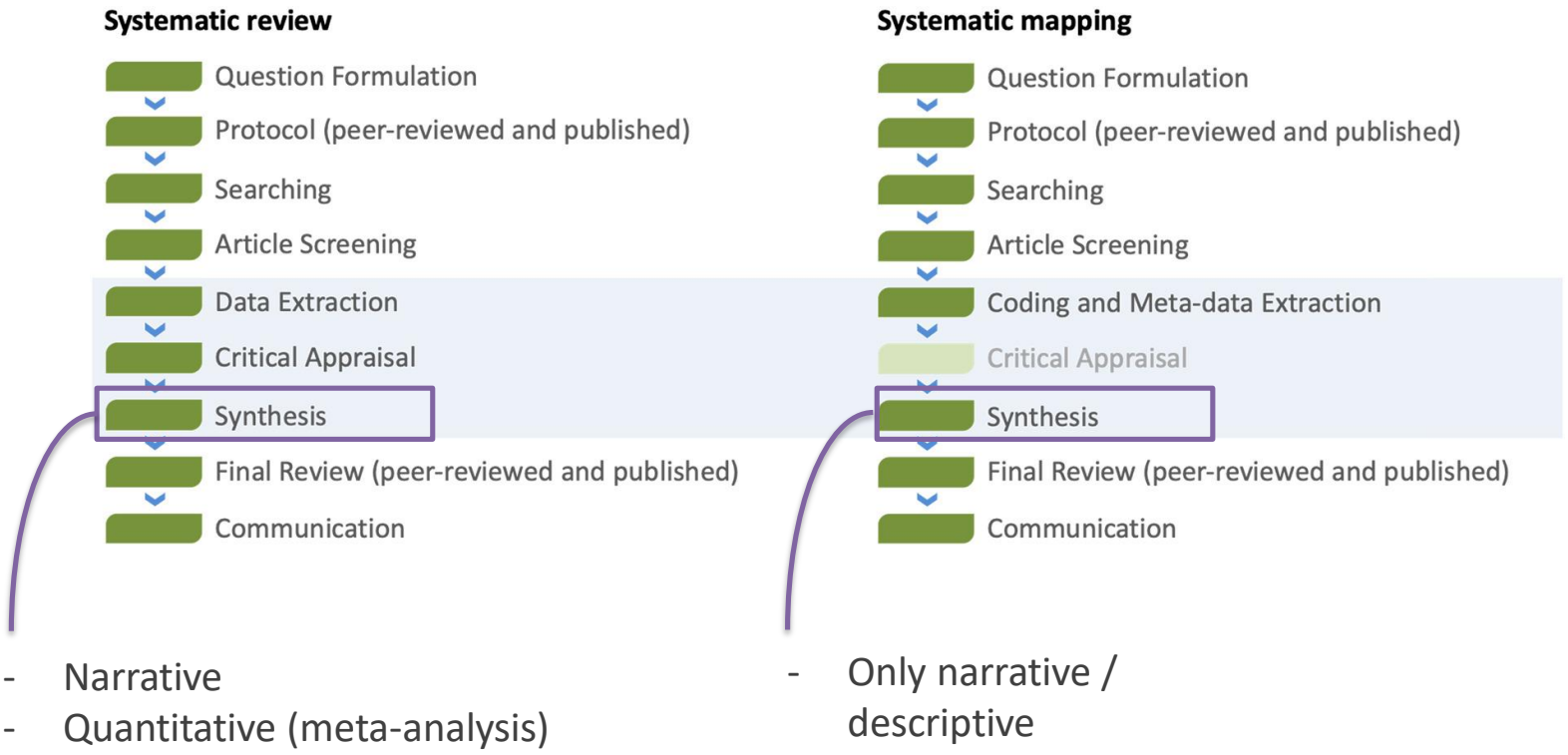
Critical appraisal

Joseph LANGRIDGE
joseph.langridge@fondationbiodiversite.fr

Screening for relevance



Comparing systematic methods

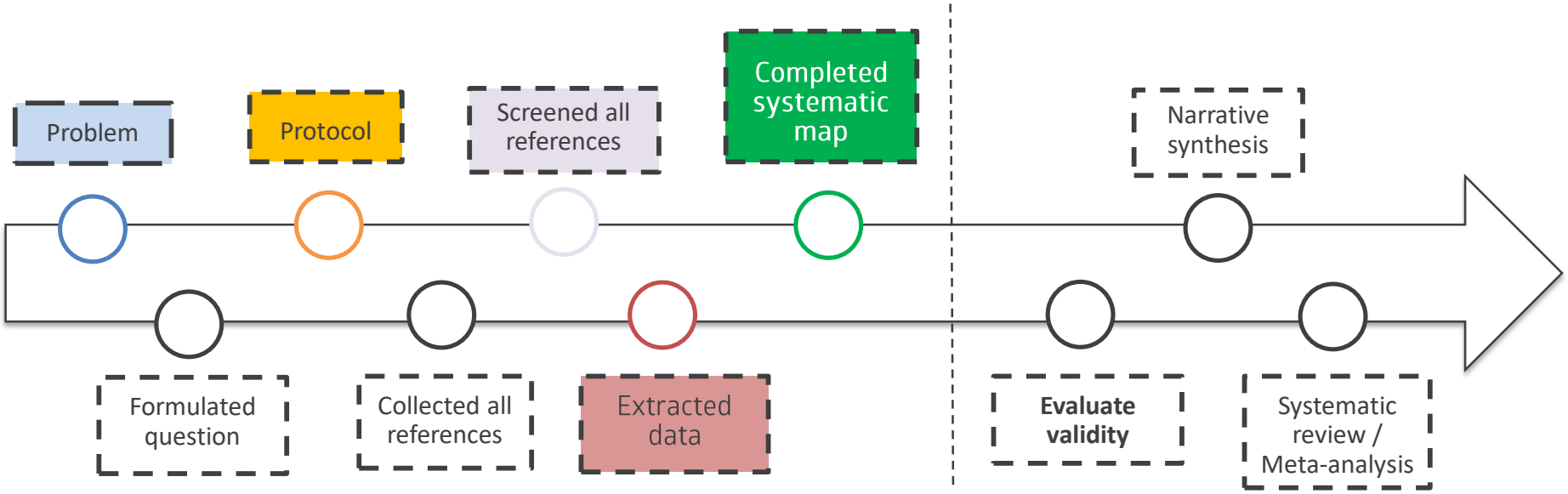


- Narrative
- Quantitative (meta-analysis)

- Only narrative / descriptive

Haddaway et al. 2020. *Campbell Systematic Reviews.* https://doi.org/10.1002/cl2.1129

Steps taken



Adapted from Bilotta *et al*. 2014. *Environmental Science & Policy*. https://doi.org/10.1016/j.envsci.2014.05.010

Why do a criticial appraisal?

- It is **rare to find scientific publications that are TRUE REPLICATES** of an initial observation or experiment. Research funding and scientific publishing encourage innovation, not verification by replication.

- It is **rare to have the means and the time to carry out 'perfect' research**, and all the more so when you are 'in the field' and not under controlled conditions...

When to do a critical appraisal?

- Critical appraisal is carried out after the various screening stages: on the **final corpus**.
  - ✓ We now know all the articles that deal with the given question (PICO/PECO).

- (Critical appraisal is carried out in a systematic review but not in a systematic map.)

- Thus, critical appraisal focuses on the ~~**quality**~~ **validity** of the studies.

- Critical appraisal is generally carried out on **studies / study units** (prior breakdown of “articles” ==> “study units”)
  - ✓ It is possible to **group together** several studies from the same publication if their experimental protocol is the same

Why do a criticial appraisal?

Extract the values (or direction of effect) from the results/outcomes obtained (whether they are significant or not)

	Comparison code	Biblio ID	Biblio author(s)	Biblio Year	Biblio title	Outcome1	Outcome2
A	Leski_2019_1	35	Leski, T; Rudawska, M	2019	Both forest reserves and ma	Increase	28.9
B	Petzold_2018_1	268	Petzold, J; Dittrich, S;	2018	Effects of forest managemer	Increase	4.63
C	Baran_2018_1	151	Baran, J; Pielech, R; B	2018	No difference in plant speci	No effect	2
D	Horvat_2017_1	302	Horvat, V; Heras, P; G	2017	Intensive forest managemer	Increase	11.41
E	Dvorak_2017_1	345	Dvorak, D; Vasutova,	2017	Macrofungal diversity patter	Increase	16.07
F	Horvat_2017b_1	370	Horvat, V; Biurrun, I;	2017	Herb layer in silver fir - beec	Increase	26.84

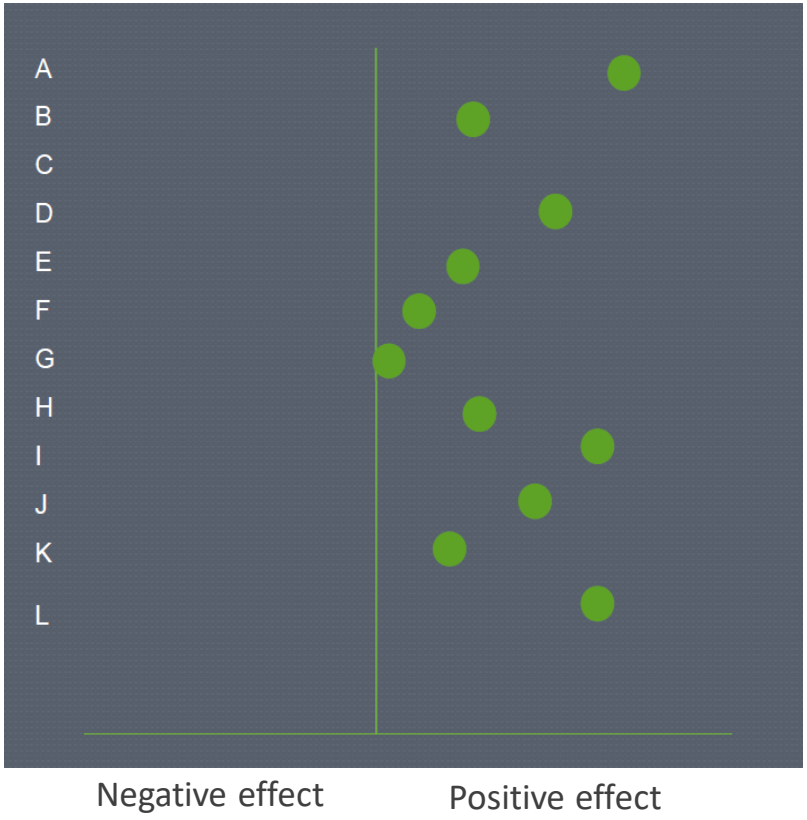
Why do a criticial appraisal?

- 12 studies (12 articles)
- PICO ou PECO

We choose an outcome (metric)

Produce a « forest plot »

Example 1: consensus?



Why do a criticial appraisal?

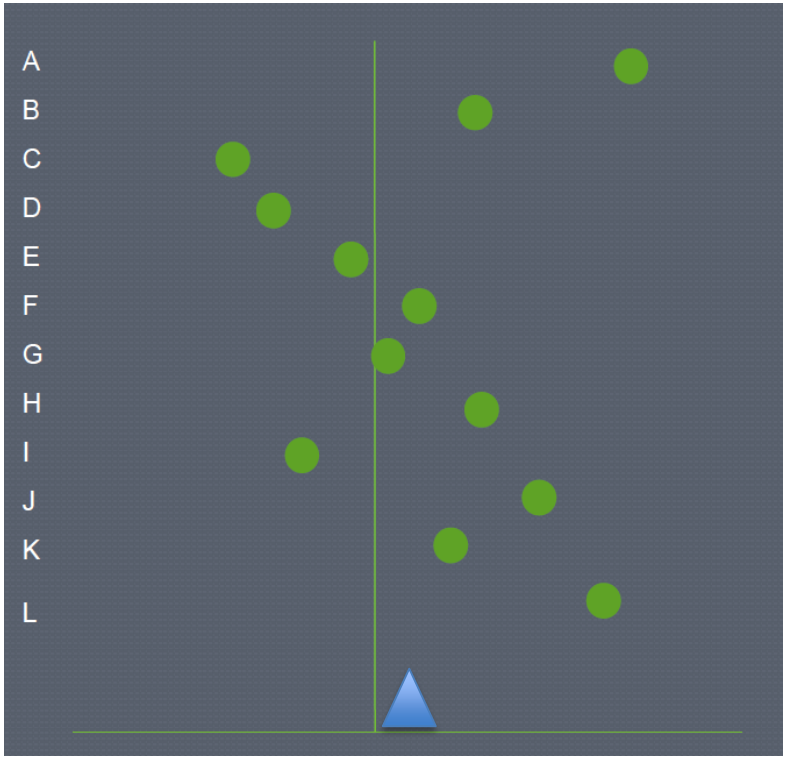
- 12 studies (12 articles)
- PICO ou PECO

We choose an outcome (metric)

Produce a « forest plot »

Hypothesis:
- Truth, reality
- Randomness
- Bias

Example 2: heterogeneity?



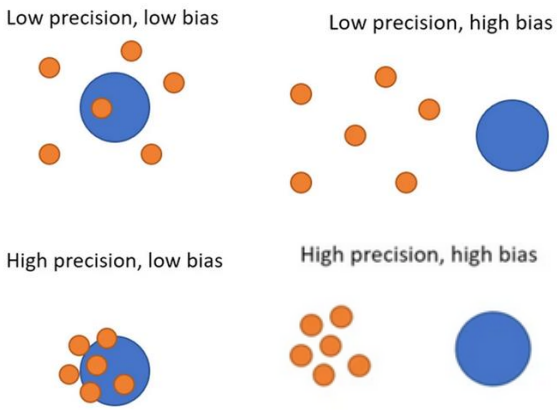
Error Vs. Bias

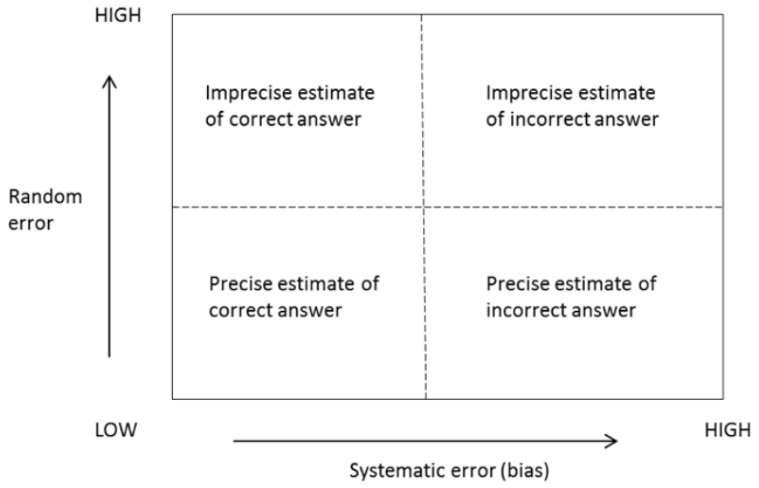
- **Random error** = imprecise (false) measurement/observation due to an accident (chance) or a temporary defect. **Unintentional.**

If measures are repeated, a good chance that it **will not be** reproduced.

- **Bias = systematic error.** It is not erased by repetition and distorts reality. It can sometimes be "voluntary".

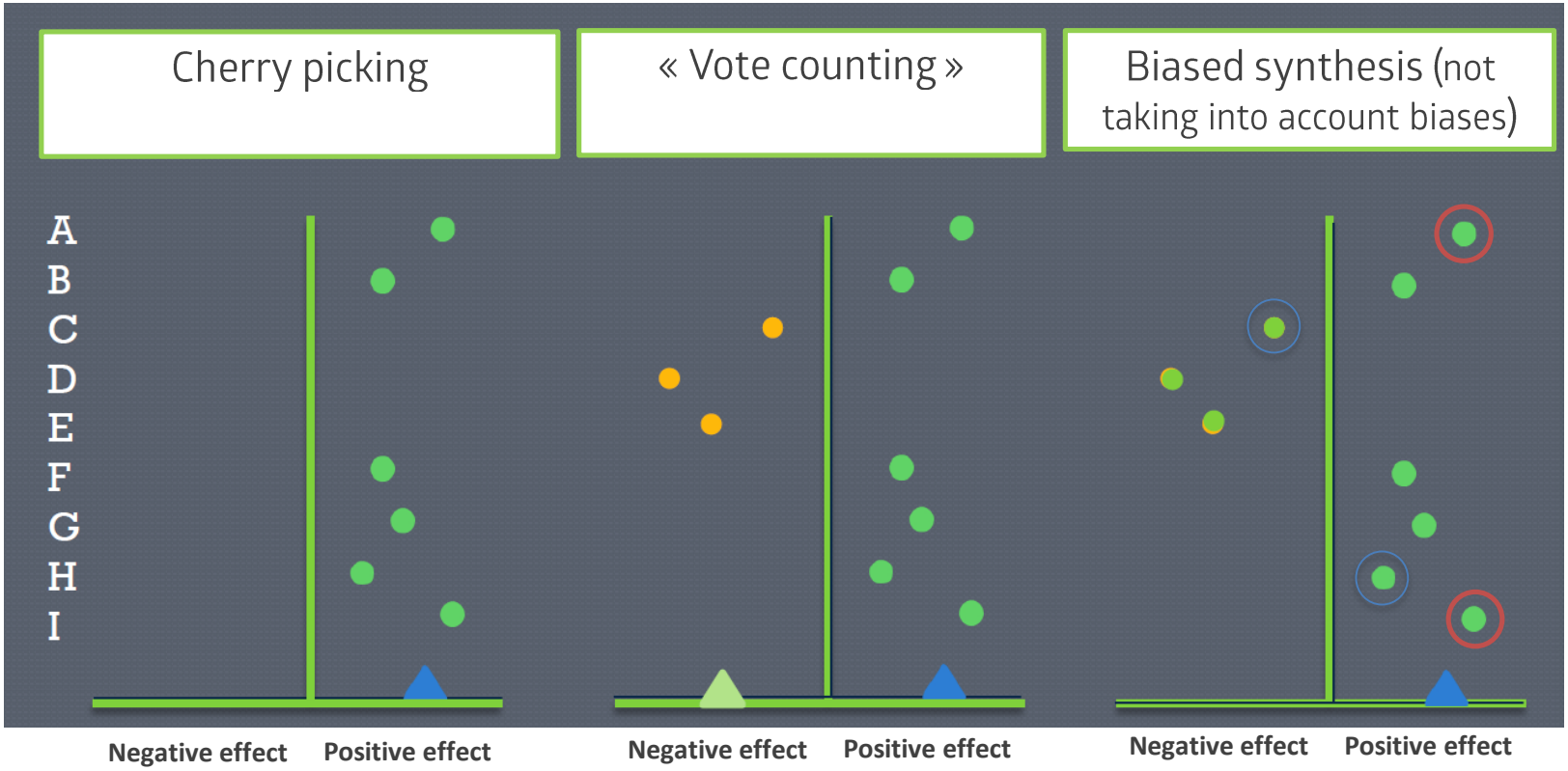
If systematic error is known but not controlled. Leading to an incorrect answer.





Frampton, *et al.* 2022. *Environmental Evidence*. https://doi.org/10.1186/s13750-022-00264-0
Konno *et al.* 2024. *Environmental Evidence*. https://doi.org/10.1186/s13750-024-00324-7

Three recurring problems



Example: biased synthesis

INSECT POPULATIONS

Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances

Roel van Klink[1,2,3*], Diana E. Bowler[1,4,5], Konstantin B. Gongalsky[6,7], Ann B. Swengel[8], Alessandro Gentile[1], Jonathan M. Chase[1,9]

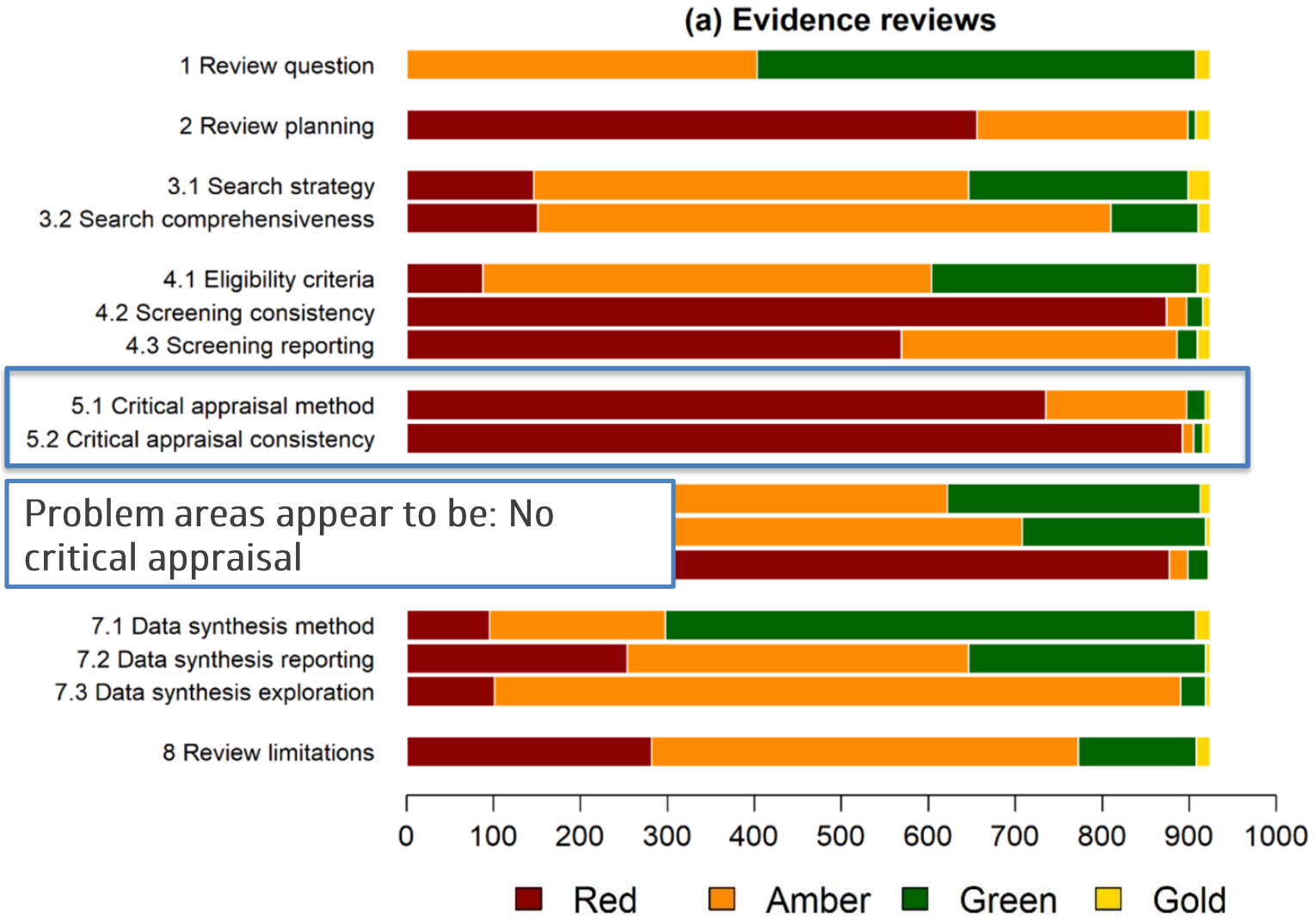
Comment on "Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances"

Marion Desquilbet[1*†], Laurence Gaume[2†], Manuela Grippa[3], Régis Céréghino[4], Jean-François Humbert[5], Jean-Marc Bonmatin[6], Pierre-André Cornillon[7], Dirk Maes[8], Hans Van Dyck[9], David Goulson[10]

https://www.science.org/doi/10.1126/science.abd8947

- No attempt was made to weight studies
- Confounding factors
  - geographic location,
  - anthropogenic impact (including farming methods and pesticide use),
  - protected status

Reliability and replicability of evidence reviews



Problem areas appear to be: No critical appraisal

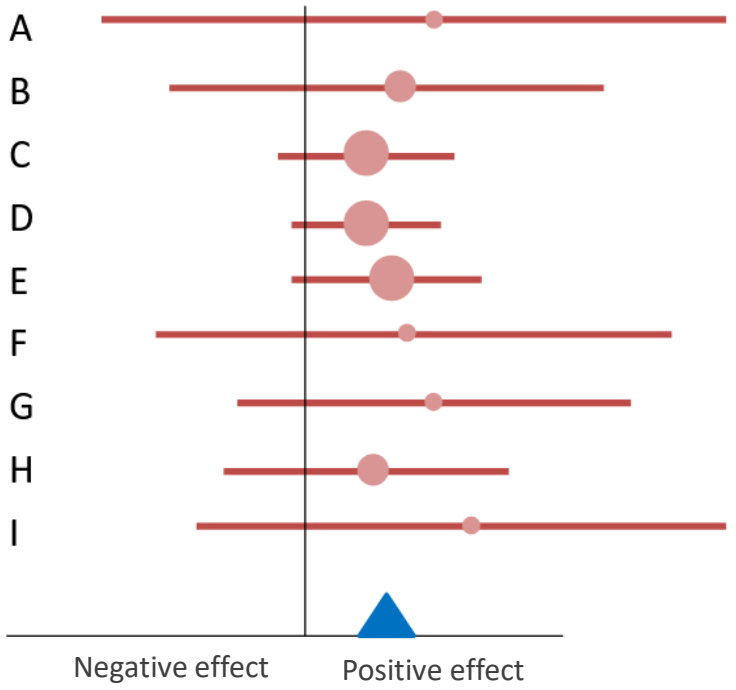
(Pullin *et al.* 2022. *Environmental Evidence*. https://doi.org/10.1186/s13750-022-00269-9)

Vote counting

Problems with « vote counting »

Significant neg effect	No signficant effect	Significant pos effect
0	9	0

Different primary studies having measured the impact of a same intervention



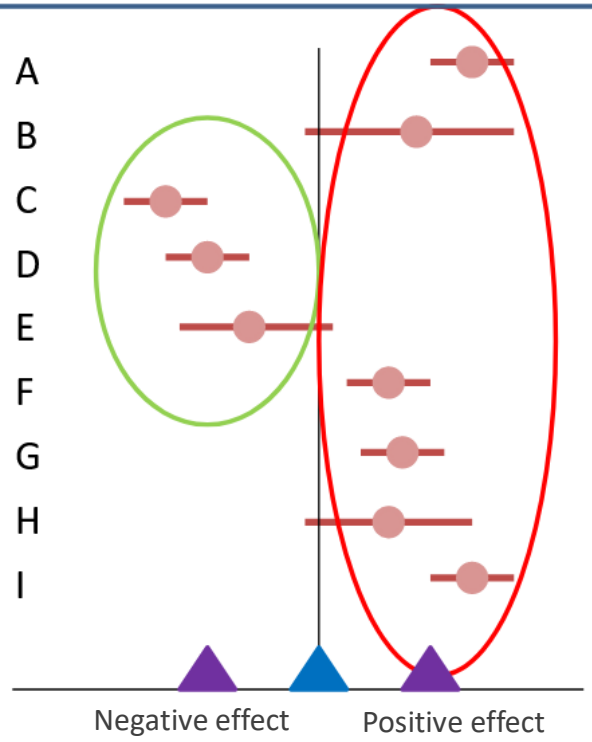
Problems with « vote counting »

Significant neg effect	No signficant effect	Significant pos effect
2	3	4

What is the reality?

Subgroup analysis

e.g. location, method, sex, species, author, season…



Why is vote-counting problematic?

Vote counting is statistically flawed because it provides only limited information about study outcomes (Gurevitch, J. *et al.* 2018. *Nature*. https://doi.org/10.1038/nature25753)

Statistical Power:

- is the probability of detecting an effect (where there is a true effect present)
- is a function of the estimated population **effect size**, the **Type I error rate** (i.e. probability of concluding that there is an effect of an intervention when there is no true relationship), and **sample size**.

Precision:

- Vote-counting does not weigh by precision. Each study is given a single "vote" in the analysis. From sampling theory, we know that **smaller samples are more likely to be further away from the population mean** (Combs et al. 2011).
- A study with a sample size of 100 would be treated the same (i.e. as a single vote) as a study with a sample size of 5 in a vote-count. **This is absurd !**

Validity

- Vote-counting treats all studies as being equally valid but this is unrealistic because different study designs can vary in their reliability (Haddaway 2017)

Grainger *et al.* 2022. Why "vote-counting" is never acceptable in evidence synthesis. Pre-print. https://doi.org/10.31219/osf.io/c49uh

Applying critical appraisal with the FEAT principles

FEAT - principles

Four basic principles of critical appraisal

Focused on an appropriate and specific validity construct (i.e. internal validity or external validity).

Extensive, capturing all aspects of the validity construct (i.e. if the construct is internal validity all the different types of bias that could arise in a given study design must be identified and assessed).

Applied – to inform the data synthesis step of the evidence synthesis in an appropriate way.

Transparent - to maximise objectivity and clarity.

cf. Frampton, et al. 2022. *Environmental Evidence*. https://doi.org/10.1186/s13750-022-00264-0

Example: What is the impact of abandoning forest management compared to continuing management on forest biodiversity?

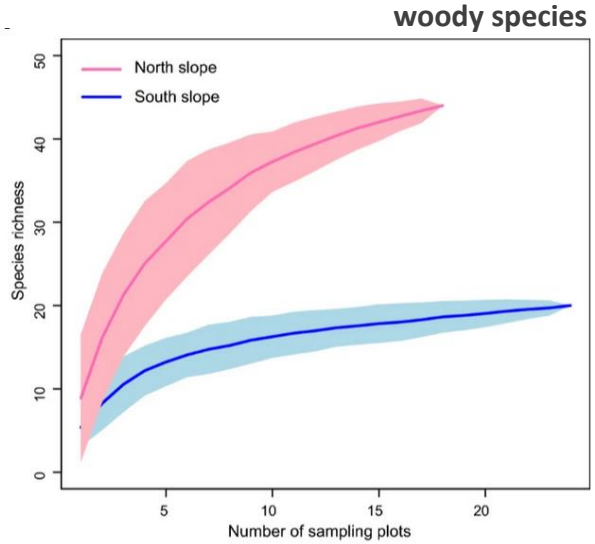
Intervention and control (C-I) sites are identical except variable of interest

Vs.

Site type
Gross successional stages
Sampling design / comparator

Langridge *et al.* 2023. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2023.110296

Example: What is the impact of abandoning forest management compared to continuing management on forest biodiversity?



Yang et al. 2020. *Scientific reports*. https://doi.org/10.1038/s41598-020-73496-0

Site type

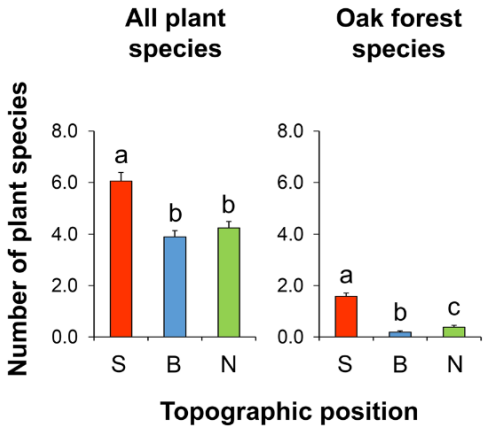


Fig. 5 Number of vascular plant species (all plant species and oak forest species) in the plots of topographic positions in dolines (S: south-facing slope, B: bottom, and N: north-facing slope) (mean ± SE). Significant differences ($p < 0.05$, based on the GLMMs with Bonferroni post hoc tests) are indicated by different lower case (a–c) letters

Bátori et al. 2023 *Annals of Forest Science.* https://doi.org/10.1186/s13595-023-01183-x

Principle:

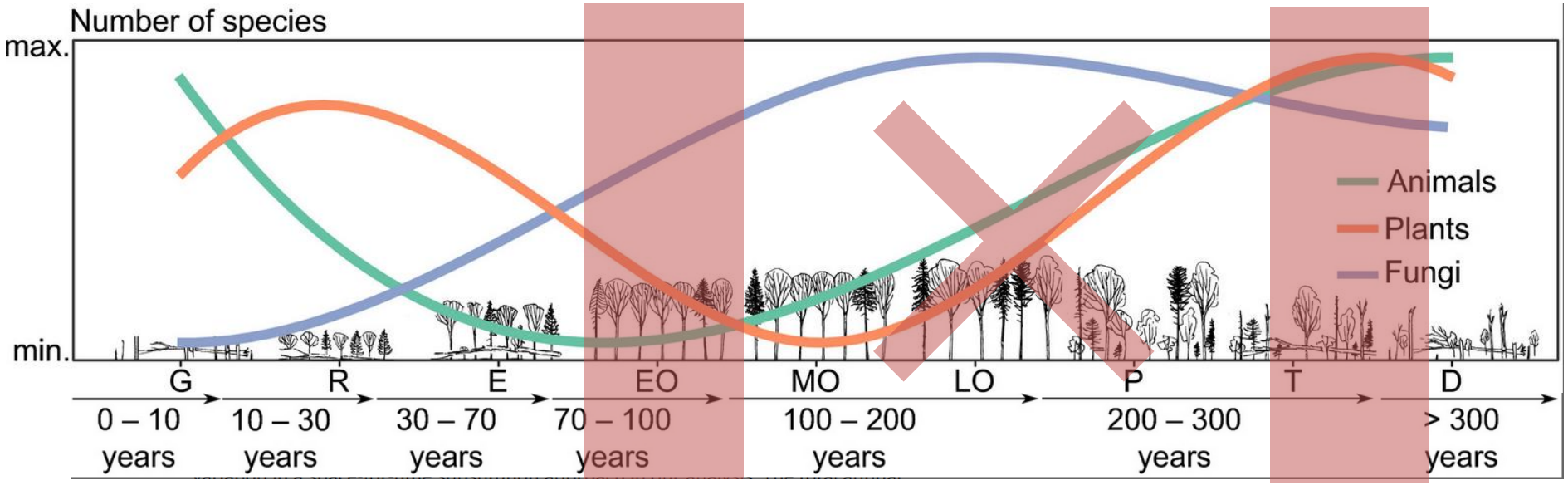
South vs North

North unmanaged vs north managed

South unmanaged vs south managed

Langridge *et al.* 2023. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2023.110296

Example: What is the impact of abandoning forest management compared to continuing management on forest biodiversity?

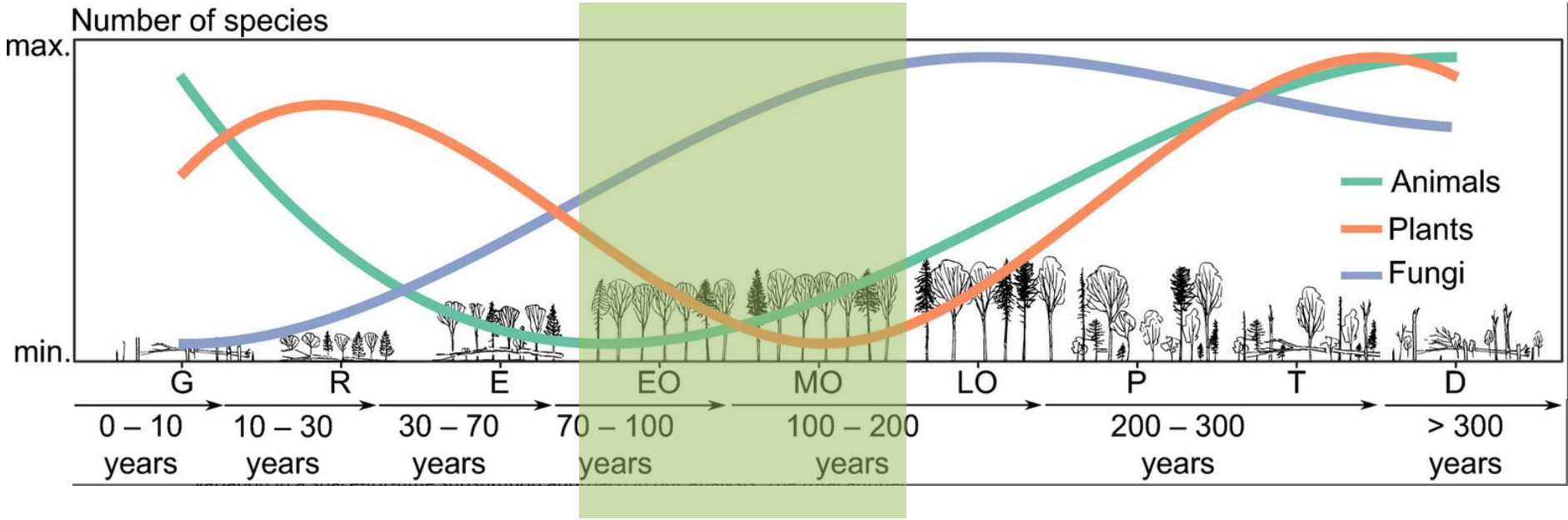


G, gap; **R**, regeneration; **E**, establishment; **EO**, early optimum; **MO**, mid-optimum; **LO**, late optimum; **P**, plenter; **T**, terminal; **D**, decay

Hilmers et al. 2018. *Journal of Applied Ecology*.
https://doi.org/10.1111/1365-2664.13238

Langridge *et al.* 2023. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2023.110296

Example: What is the impact of abandoning forest management compared to continuing management on forest biodiversity?



Respecting successional stages

“Gross successional” stages

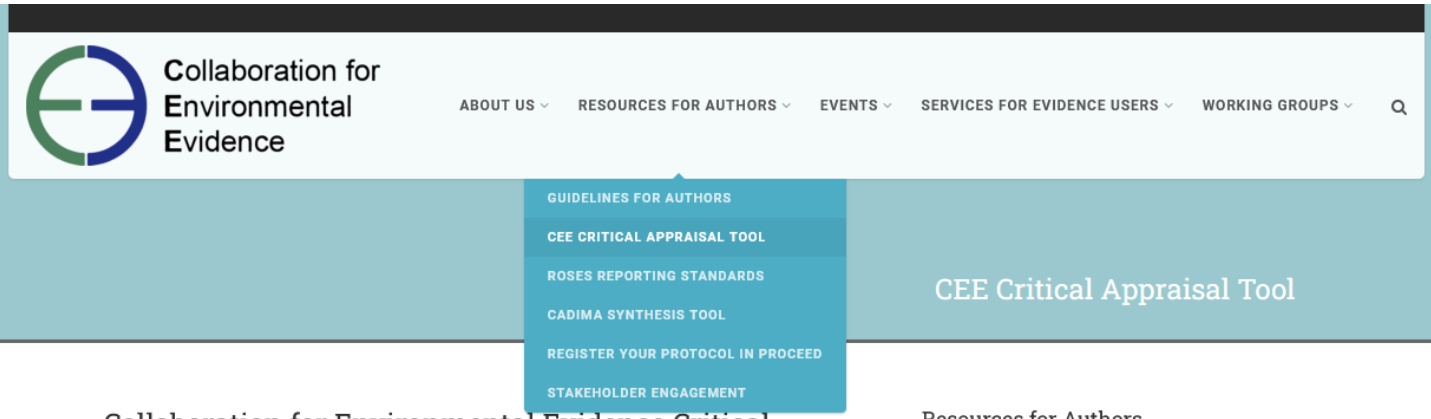
mid managed vs mid unmanaged

Sampling design / comparator

Langridge *et al.* 2023. Biological Conservation. https://doi.org/10.1016/j.biocon.2023.110296

Introducing
CEE Critical Appraisal Tool

CEE : Critical Appraisal Tool



Collaboration for Environmental Evidence Critical Appraisal Tool Version 0.3 (Prototype)

Last updated: 24 October 2021

PROJECT SUMMARY

We are currently developing a critical appraisal tool for evaluating 'risk of bias' (or threats to internal validity) of primary studies assessing effectiveness of interventions or impacts of exposures in environmental management. There are well-known, widely applied risk of bias assessment tools in the health sector known as 'RoB 2' and 'ROBINS-I' (www.riskofbias.info), but there are currently no such critical appraisal tools in environmental management. Here we provide a third draft of the tool. The tool is still under development and requires initial testing, but it may help environmental evidence synthesists conduct critical appraisal. In the latest version, more descriptions are provided in Part A, and revisions and clarifications are made in Part B. We have merged two risk-of-bias criteria and so the total number of risk-of-bias criteria is reduced to seven in version 0.3.

If you are not familiar with critical appraisal step, we recommend reading Sections 3.5 (www.environmentalevidence.org/guidelines/section-3) and 8 (www.environmentalevidence.org/guidelines/section-8) of *CEE Guidelines* before you download the

Resources for Authors

Guidelines for Authors

- Aims and Scope
- Table of Contents
- Updates and Corrections
- 1. Process Summary
- 2. Need for Evidence, Synthesis Type and Review Team
- 3. Planning a CEE Evidence Synthesis
- 4. Writing and Registering a Protocol
- 5. Conducting a Search
- 6. Eligibility Screening
- 7. Data Coding and Data Extraction
- 8. Critical appraisal of study validity (SRs)
- 9. Data Synthesis
- 10. Interpreting findings and reporting

Scope of tool

Within the scope

If all of the followings apply

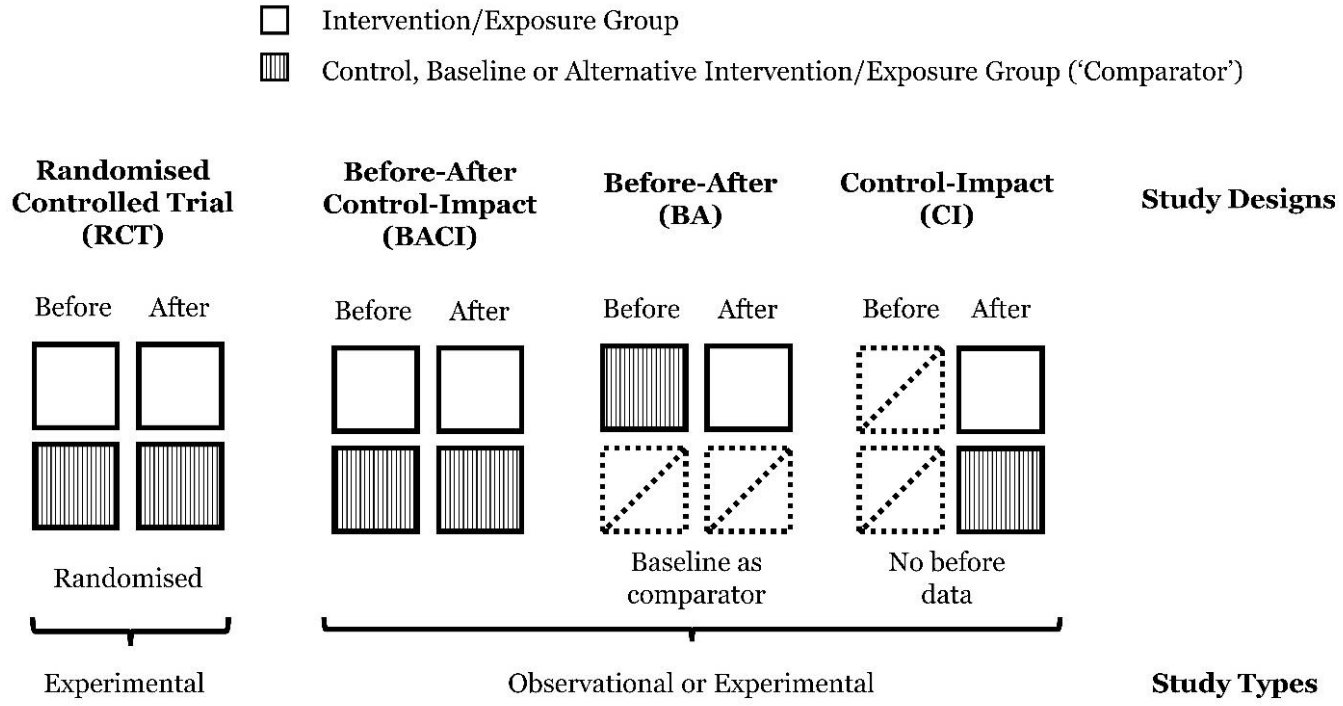
- Review question justifiably relates to environmental management (policy or practice)
- Reviewing evidence on impact of exposure or effectiveness of intervention
- Interest is quantitatively measured outcomes

Beyond the scope

If any of the followings applies

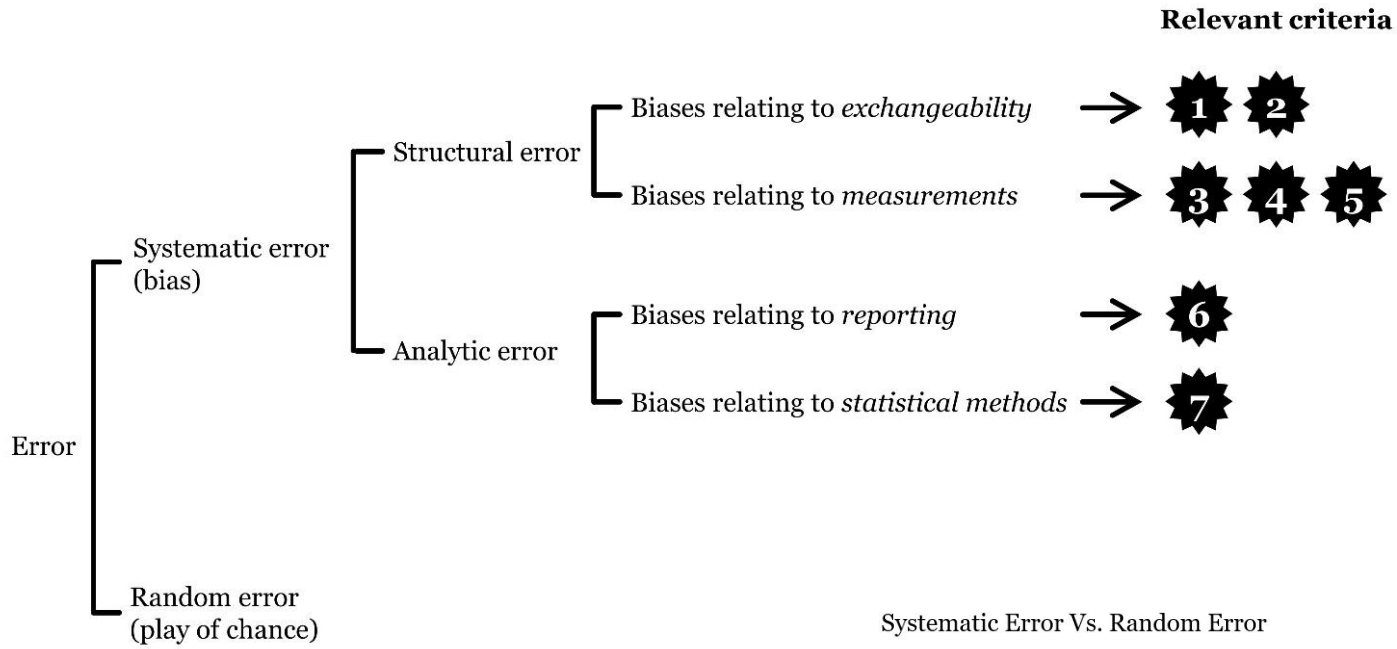
- Reviewing medical research involving human subjects, tissues, or personal data (including physiological, biomechanical, psychological research)
- Reviewing purely laboratory-based biological research (e.g., *in vitro* or *in vivo* experiments, genome sequencing)
- Reviewing qualitative evidence

Assumptions of the tool

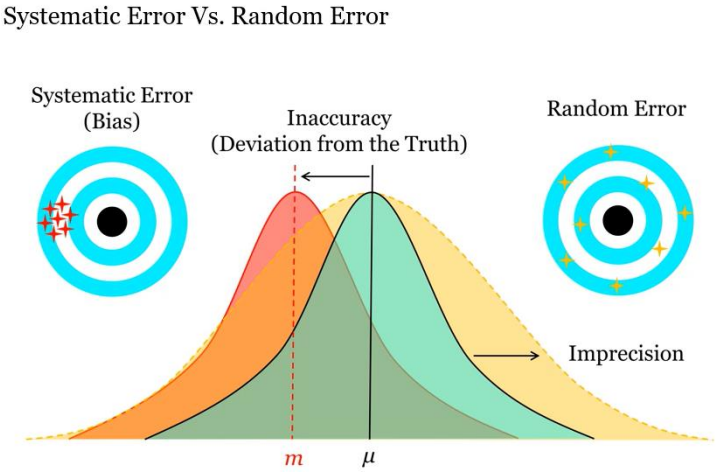


- ✓ The tool does not assume that study types are biased.
- ✓ The tool does not assume the differences in study design alone affect precision.

Organisational Schema for Errors



Structural error: bias relating to measurement of intervention, exposure or outcome, and bias relating to exchangeability. Exchangeability refers to independence between the outcome and the observed intervention.

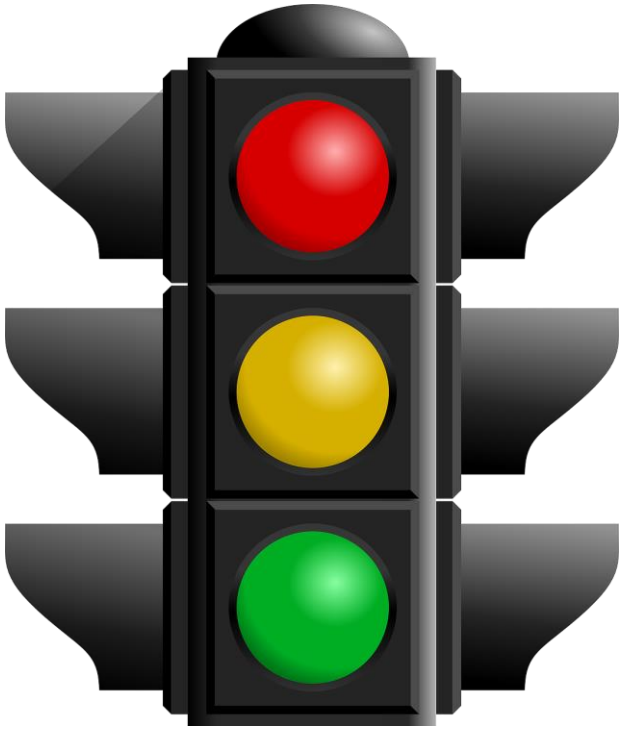


Risk-of-Bias Judgement

Within Individual Criteria

Three levels:

- Low risk of bias (**Low**)
- Medium risk of bias (**Med**)
- High risk of bias (**High**)



Internal validity: what should I check?

Risk of bias class	Summary
1. Bias due to confounding factors	**"Risk of confounding biases"** in the CEE tool. These biases arise due to one or more uncontrolled variables (confounders) that influence both the exposure and the outcome.
2. Bias in selection of subjects/areas into the study	**"Risk of post-intervention/exposure selection biases"** in the CEE tool. These biases arise when some eligible subjects are excluded in a way that leads to a false association between the exposure and outcome.
2a. Bias due to missing data (attrition bias)	Bias due to missing data can be considered as a type of selection bias; **"Risk of post-intervention/exposure selection biases"** (i.e. bias class 2 above). This can arise when follow up data of subjects - that are initially included in the study - are not fully available for inclusion in the analysis. Thus, an imbalance in the amount of missing data between the exposure and comparator groups (differential missingness)
3. Bias due to misclassification of the exposure	**"Risk of misclassified comparison biases"** in the CEE tool. These biases arise from misclassification or mismeasurement of the exposure and/or comparator which leads to a misrepresentation of the association between the exposure and the outcome (also known as measurement bias or information bias).
4. Bias due to deviation from the planned exposure (intervention) in experimental studies	**"Risk of performance biases"** in the CEE tool. These biases arise from alteration of the planned exposure or comparator treatment procedure(s) of interest after the start of the exposure.

Frampton, et al. 2022. *Environmental Evidence*. https://doi.org/10.1186/s13750-022-00264-0
Konno *et al.* 2024. *Environmental Evidence*. https://doi.org/10.1186/s13750-024-00324-7

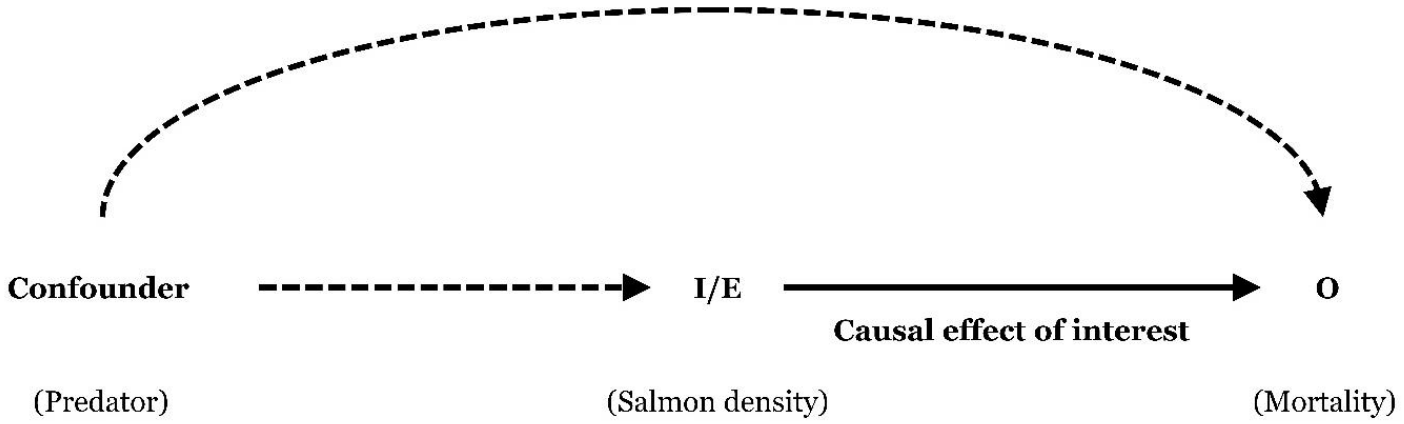
Internal validity: what should I check?

Risk of bias class	Summary
5. Bias in measurement of outcomes	**"Risk of detection biases"** in the CEE tool. This can arise from systematic differences in measurements of outcomes (also known as measurement bias). Systematic errors in measurement of outcomes may occur if outcome data are determined differently between the exposure and comparator groups. This could be intentional (e.g. influence of desire to obtain a certain direction of effect) or unintentional (e.g. due to cognitive bias or human errors).
6. Bias in selection of the reported results	**"Risk of outcome reporting biases"** in the CEE tool. These are biases arising from selective reporting of study findings. Selective reporting may appear at three different levels: (i) presentation of selected findings from multiple measurements; (ii) presentation of results for selected subgroups or subpopulations of the planned analysis population; and (iii) presentation of selective findings from multiple analyses.
7. Bias due to an inappropriate statistical analysis approach	**"Risk of outcome assessment biases"** in the CEE tool. These are due to errors in statistical methods applied within the individual studies included in a systematic review. Four main areas: (i) data analysts' awareness of the exposure or comparator received by study subjects or areas (blinding of data analysts could mitigate the risk of bias); (ii) errors in applied descriptive statistics (e.g. miscalculation of sample sizes, means, or variances, including pseudoreplication); (iii) errors in applied inferential statistics (including flawed null hypothesis testing, estimation, or coding); (iv) use of inappropriate statistical tests or violation of assumptions required by tests (e.g. criteria for normality and equal variances are not satisfied)

Frampton, et al. 2022. *Environmental Evidence*. https://doi.org/10.1186/s13750-022-00264-0
Konno *et al.* 2024. *Environmental Evidence*. https://doi.org/10.1186/s13750-024-00324-7

Example: Criterion 1

Risk of Confounding Biases

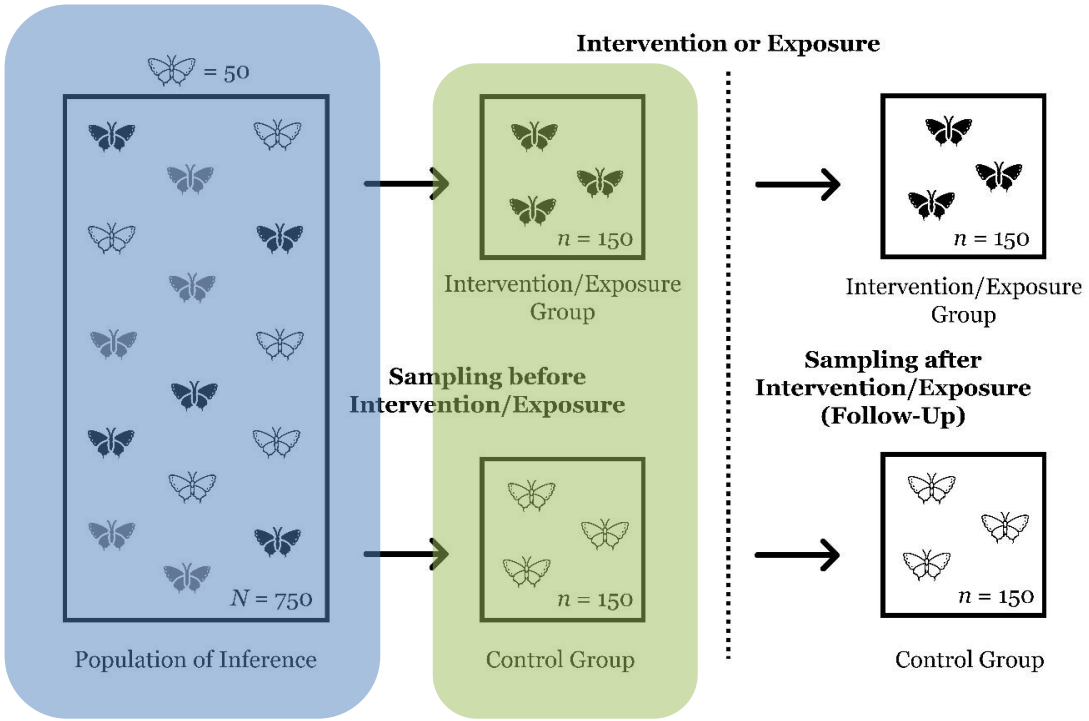


Understand the effect of salomon denisty on their survival?

Presence of predators → affect rate of survival (O) thus density (I/E).

Example: Criterion 1

Risk of Confounding Biases



Need to understand variability

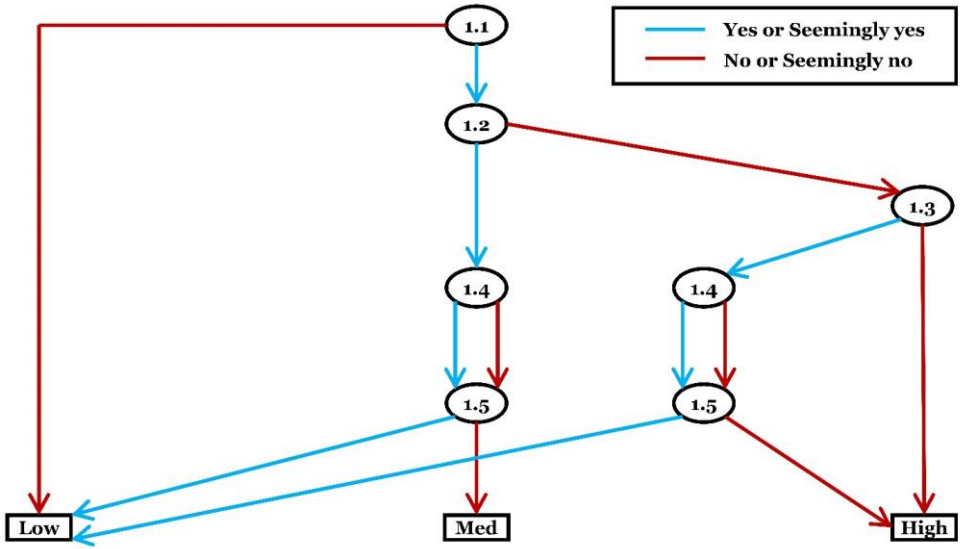
e.g. morphological differences (covariate) may explain the effect

⇩

Stratified sampling
Divided into sub-populations

Example: Criterion 1

Risk of Confounding Biases



1.1. Is confounding possible?

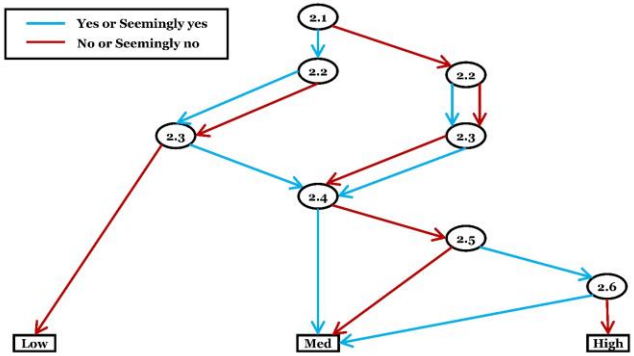
1.2. Are the potential confounders controlled for?

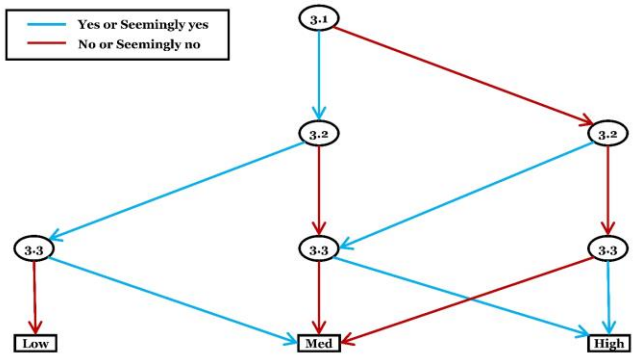
1.3. Is there any justification for not controlling for the potential confounders?

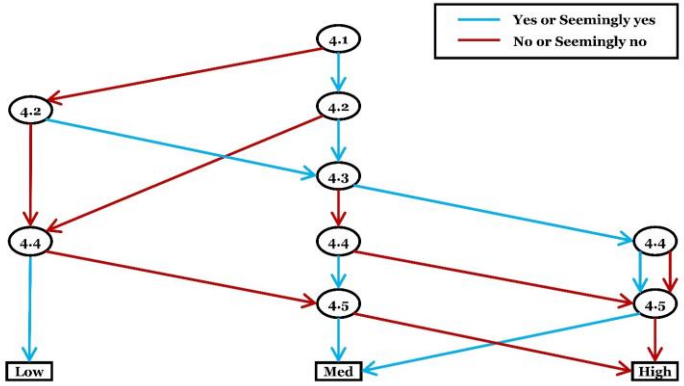
1.4. Are the potential confounders likely to be measured accurately and precisely?

1.5. Are the analyses of the effect appropriate?

Decisions trees for each criterion







Criterion 1: Risk of confounding biases

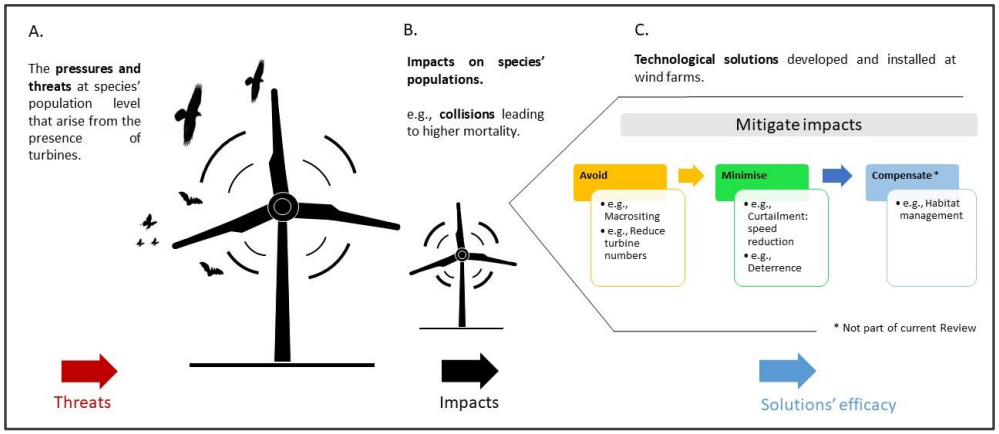
Category	Checklist questions	Response option	Answer
General (please answer)	1.1. Is it possible for the impact of the exposure or the effectiveness of the intervention to be confounded in this study?	Yes	
		Seemingly yes	
		Seemingly no	
		No	
Conditional (answer if Y/SY to 1.1, otherwise select 'Not applicable')	1.2. Did the author(s) control for all the potential confounders?	Yes	
		Seemingly yes	
		Seemingly no	
		No	
		Unclear (No)	
		Not applicable	
Conditional (answer if N/SN/Unclear to 1.2, otherwise select 'Not applicable')	1.3. Is there any justifiable reason for not controlling for all the potential confounders (so that omission of some of the potential confounders is unlikely to influence the assessment of the effectiveness or impact)?	Yes	
		Seemingly yes	
		Seemingly no	
		No	
		Not applicable	

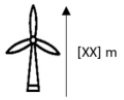
Citation | Criterion1 | Criterion2 | Criterion3 | Criterion4 | Criterion5 | Criterion6 | Criterion7 | Overall | All_decisions

Konno K, Livoreil B, Pullin AS. 2021. Collaboration for Environmental Evidence Critical Appraisal Tool version 0.3. https://environmentalevidence.org/cee-critical-appraisal-tool/

How to carry it out?

Quinard *et al.* The effectiveness of existing solutions to mitigate impacts of onshore wind farms on flying vertebrates and invertebrates: a Rapid review. *Environmental evidence*. [in prep]



Type of comparative study desgin	Illustrated example		Other examples of comparative study designs
Raising cut-in speed threshold	Wind speed in m/s : higher cut-in speed 	Wind speed in m/s : lower cut-in speed 	• Targeted cut-in speed • Feathering • Cut-in speed and simultaneous acoustic deterrence
Acoustic deterrence			• Radar deterrence • UV light deterrence • Radar and acoustic deterrence simultaneously
Turbine size	 [XX] m	 [XX] m	• Repowering
Surface painting (including the blades)			• Surface aspect
Micro-siting			• Macro-siting • Elimination of attraction factors • Elimination of attraction factors: light

CEE appraisal tool applied

Quinard *et al.* The effectiveness of existing solutions to mitigate impacts of onshore wind farms on flying vertebrates and invertebrates: A Rapid Review. [in prep]

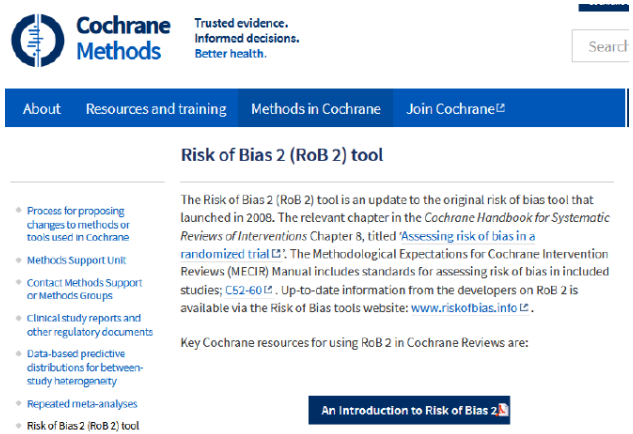
Risque de biais	Question	Low	High	Medium	Unclear or unknown
Internal validity bias	Does the intervention measure take place *in situ* (on a site equipped with wind turbines)?	Yes, or presumably yes	No (*ex-situ* studies), or presumably no	NA	Unclear or Unknown
Confounding factors bias	Are there any potential confounding factors (cf. sheet 2) that could influence the intervention and/or the outcome? If so, have the authors identified, analyzed/controlled, and taken them into account in their analysis?	No, or presumably no	Yes, or presumably yes	Yes, but poorly controlled for	Unclear or Unknown
Misclassified comparison bias (only observational)	Are the exposure/intervention and comparison groups sufficiently well defined?	Yes	No, or presumably no	NA	Unclear or Unknown

CEE appraisal tool applied

Quinard *et al.* The effectiveness of existing solutions to mitigate impacts of onshore wind farms on flying vertebrates and invertebrates: A Rapid Review. [in prep]

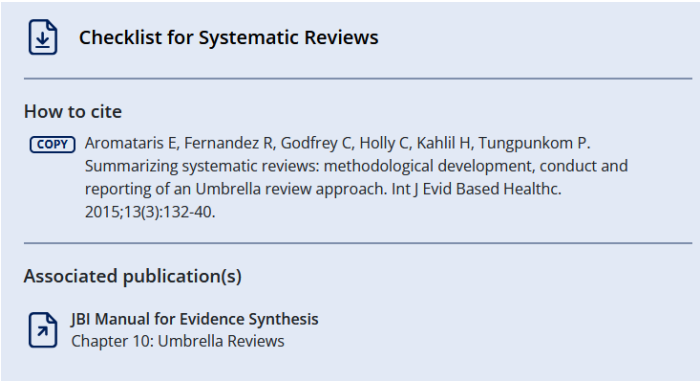
				Validité externe	Facteurs confondants	Sélection post-exposition			Classement incorrect (uniquement études observationnelles)	Performance (uniquement études expérimentales)			
				L'exposition/ intervention a-t-elle lieu in-situ (sur site équipé d'éoliennes)?	Existe-t-il de potentiels facteurs confondants (voir feuille 2) pouvant influencer l'intervention et/ou le résultat ? Si oui, les auteurs les ont-ils identifié, puis analysé/contrôlé, et les ont-ils pris en compte dans leur analyse ?	La sélection des sujets ou des zones après l'intervention ou l'exposition était-elle aléatoire ou systématique, et pourrait-on supposer l'interchangeabilité entre les groupes avant et après?	Le groupe assigné aux sujets ou aux zones (type d'intervention/contrôle) était-il caché aux expérimentateurs ?	Y avait-il un nombre différent de données manquantes entre les groupes exposés et les groupes témoins pendant l'étude ou l'analyse ?	Les groupes d'exposition/intervention et de comparaison sont-ils suffisamment bien définis ?	Y a-t-il eu des modifications des traitements d'intervention/exposition ou de contrôle d'intérêt qui pourraient avoir un impact sur l'efficacité de l'intervention ou l'impact de l'exposition ?	Les tailles d'échantillon de ces traitements altérés étaient-ils déséquilibrés entre les groupes d'intervention ou d'exposition ou ces traitements altérés étaient-ils pris en compte de façon incorrecte, ce qui pourrait avoir influencé l'estimation de l'impact ou de l'efficacité ?	Dans le cas d'une évaluation de la mortalité, un test de persistance (facteur de correction) a-t-il été réalisé ? Si oui, prend-il en compte : la taille des carcasses, les mesures pour chaque turbine séparémment ? De même, un test de détection avec un contrôle pour les différences spécifiques au site a-t-il été réalisé ?	La variation de l'efficacité entre les observateurs et au fil du temps (facteur de correction) a-t-elle était évaluée et utilisée?
Article		**Evaluateur**	**CrossCheck**										
May_2017	Do birds in flig	**AQ**		High	Low	NA	NA	Unknown	NA	Low	NA	NA	NA
Cooper_2020	Bat impact mir	**AQ**	**LD**	Low	High	Low	Unknown	Unknown	NA	High	High	Medium	Unknown
Smales_2013	A description c	**AQ**		Low	Low	NA	NA	NA	NA	NA	NA	Medium	Unknown
Nicholls_2009	The aversive ef	**AQ**		High	High	NA	Low	Unknown	NA	Low	NA	NA	NA
Pescador_2019	Effectiveness o	**Pre-test**		Low	Medium	Unknown	Unknown	Unknown	NA	Low	NA	Unknown	Unknown
Bienz_2016	Surface texture	**AQ**		High	High	Low	High	Unknown	NA	Low	NA	NA	NA
Smallwood_2020	Effects of wind	**AQ**	**JL**	Low	Medium	Unknown	Unknown	Unknown	NA	Low	NA	Low	Unknown
Gorresen_2015	Dim ultraviolet	**Test**		High	Low	Low	Unknown	Low	NA	Low	NA	NA	NA
Stokke_2020	Effect of tower	**AQ**	**LD**	Low	Low	Unknown	High	Unknown	NA	Low	NA	High	High
Ferri_2016	Bats in a medit	**AQ**		Low	Low	Unknown	Low	Unknown	NA	Low	NA	NA	NA
Weaver_2020	Ultrasonic acou	**AQ**	**JL**	Low	Medium	Low	Unknown	High	NA	High	High	Medium	High

Available TOOLS

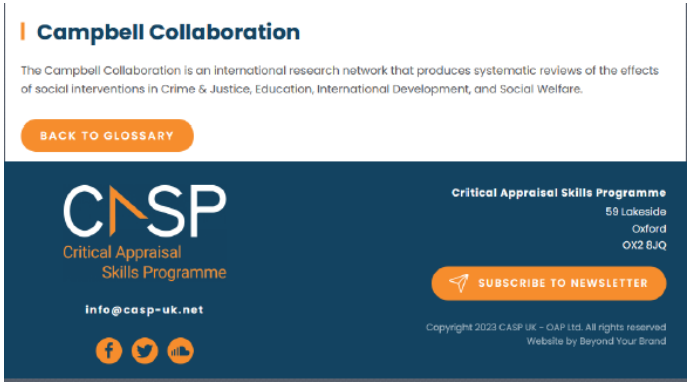


https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials

https://environmentalevidence.org/cee-critical-appraisal-tool/



https://jbi.global/critical-appraisal-tools



https://casp-uk.net/casp-tools-checklists/