Full title: Elevating the importance of Risk of Bias assessment for ecology and evolution

Antica Culina*, Ruder Boskovic Institute, Zagreb, Croatia (https://orcid.org/0000-0003-2910-8085), aculina@irb.hr

Dugald Foster, Centre for Ecology and Conservation, University of Exeter, Penryn, UK (https://orcid.org/0000-0002-0727-830X)

Matt Grainger, Knowledge Synthesis Department, P.O. Box 5685 Torgarden, 7485, Trondheim, Norway (https://orcid.org/0000-0001-8426-6495)

Rose E. O'Dea, School of Agriculture, Food and Ecosystem Sciences, University of Melbourne, Melbourne, Australia (https://orcid.org/0000-0001-8177-5075)

Oliver L. Pescott, UK Centre for Ecology & Hydrology, Wallingford, UK (https://orcid.org/0000-0002-0685-8046)

Alfredo Sánchez-Tójar, Department of Evolutionary Biology, Bielefeld University, Germany (https://orcid.org/0000-0002-2886-0649)

Robin J. Boyd, UK Centre for Ecology & Hydrology, Wallingford, UK (orcid.org/0000-0002-7973-9865)

Julia Koricheva, Department of Biological Sciences, Royal Holloway University of London, Egham, Surrey, UK (https://orcid.org/0000-0002-9033-0171)

Shinichi Nakagawa, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada (https://orcid.org/0000-0002-7765-5182)

Gavin Stewart, University of Newcastle, UK (https://orcid.org/0000-0001-5684-1544)

Statement of Authorship

Conceptualization: AC; Data curation: AC; Funding acquisition: AC; Investigation: AC, MG, REO, OP, AST; Methodology: all authors; Project administration: AC; Software: AC, AST; Supervision: AC; Visualization: AC; Writing – original draft: AC; Writing – review & editing: all authors.

Data accessibility statement

Data and scripts used in this work are available at Zenodo: Culina, A., Sánchez-Tójar, A., Pescott, O., O'Dea, R., & Grainger, M. (2025). Datasets and codes for 'Elevating the importance of Risk of Bias assessment for ecology and evolution' [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15630232

Keywords: Evidence-base, Study quality, Critical appraisal, Methodological transparency, Research synthesis, Reproducibility, Bias assessment, Research integrity, Meta-research, Best practices

Correspondence to: Antica Culina, +385 919187450, Institute Ruder Boskovic, Bijenička cesta 54, 10000 Zagreb, Croatia, aculina@irb.hr

Abstract

Systematic reviews and meta-analyses are key evidence synthesis methods informing research and policy. An assessment of the Risk of Bias (RoB) in included studies is normally considered an essential component of these. However, RoB assessment is rare in ecology and evolutionary biology (EEB), and tools from other fields are seldom adopted. To identify reasons for this limited uptake, we surveyed 232 ecologists and evolutionary biologists with evidence synthesis experience. Only 28 were familiar with the RoB concept, and 46 confused it with publication bias. Just 10 had conducted a RoB assessment, most of whom found it challenging. We also reviewed the guidelines of 275 journals publishing EEB research. Out of 209 journals that likely solicit evidence synthesis, only five referenced guidelines for conducting evidence synthesis, and thus including of RoB assessment. To elevate the status of RoB assessment we recommend: (1) recognising RoB as essential to reliable evidence synthesis, with journals promoting clear guidelines and policies; (2) improving reporting standards for primary studies to enable RoB evaluation; and (3) developing RoB tools that accommodate EEB research and training researchers in their implementation. To start, we propose four RoB questions researchers should consider when conducting evidence synthesis or designing primary studies.

Introduction

Do search costs of mate choice enhance speciation by sexual selection? To what degree do trans-generational effects on life histories, such as maternal effects, impact population dynamics? These and other questions in ecology and evolution can be investigated by conducting experimental or observational studies (i.e., primary studies). However, the findings of primary studies are often limited to a particular, often narrow context, and generalization frequently requires a synthetic view across primary studies. Further, primary studies will exhibit some combination of random error, systematic bias, and contextual variation that must be assessed before findings can be generalised. Evidence synthesis integrates information across primary studies, and thus often across taxonomic, geographical, and other scales, if relevant. Information on a topic of interest is systematically searched for and reviewed, commonly via a systematic review, and often subsequently quantitatively summarised using meta-analysis. Evidence synthesis is becoming essential to inform research and policy development, and communicate scientific findings to the public (Cooke et al. 2017). A robust evidence synthesis can also be a starting point for research or grant applications (Nakagawa et al. 2020; Grainger et al. 2020). Thus, it is of the utmost importance that systematic reviews and meta-analyses are done and reported according to the highest standards. One such key standard is the assessment of the reliability of the included primary studies, but this is often neglected in ecology and evolutionary biology despite findings of low reliability in these fields (Else 2024). Assessing the reliability of the included primary studies is key because the validity of the conclusions of a synthesis will, among others, depend on the validity of the results of such studies. Any "systematic deviation from the truth", that is, any systematic bias (e.g., Wang et al. 2022) will lead to invalid conclusions of a primary study, but also of a synthesis of such studies.

Systematic reviews and meta-analyses in ecology and evolutionary biology have adopted more rigorous methodological and reporting standards over time (Koricheva and Gurevitch, 2014, Lodi et al. 2021). Many of these were taken and adjusted from medicine, a field where reliable systematic reviews and meta-analyses are essential because of their implications for human health interventions. Examples of such standards are the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines and checklist (Moher et al. 2009; Page et al. 2021) and ROSES (RepOrting standards for Systematic Evidence Syntheses in environmental research: Haddaway et al. 2018). The former was recently expanded to ecology and evolutionary biology as PRISMA-EcoEvo (O'Dea et al. 2021). Another relevant standard is the methodological approach to testing for publication bias (Nakagawa et al. 2021). These two standards are applied after the evidence has already been included in the meta-analysis. However, what is missing is a step that evaluates the evidence before summarising it, that is, assessing the internal validity of the primary studies included. The potential threats to the internal validity of a study are often termed the "Risk of Bias" (RoB). We note that the general phrase "risk of bias" can sometimes describe other aspects related to bias, usually a bias within the overall pool of studies, such as geographic (e.g., Martin et al. 2012), language (e.g., Nunes et al, 2021), taxonomic (e.g,. Troudet et al. 2017), or publication bias (e.g., Nakagawa et al. 2022). Here, we focus on within-study RoB and argue that assessing the RoB in each included primary study should be a key standard in evidence synthesis. Moreover, beyond the context of evidence synthesis, researchers can use RoB tools to help minimise bias when designing and conducting primary studies.

With this contribution, we aim to raise awareness of and promote the use of RoB assessments in ecology and evolutionary biology by: (1) highlighting the importance of performing such assessments in evidence synthesis; (2) surveying evidence synthesists and journals to understand the main reasons for the low prevalence of RoB assessments in these fields; (3)

suggesting key actions to promote the application of RoB assessment; and, (4) suggesting a set of questions for an easy, yet informative, RoB assessment, serving as a starting point for transitioning towards higher standards for ecological and evolutionary evidence synthesis. Importantly, this set of questions can also help decrease RoB when designing primary research. Applying the RoB approach to both the design and evaluation of primary studies can provide critical insights into the limitations of bias reduction in ecology and evolutionary biology, and, by extension, into the strength of the inferences we can draw from ecological and evolutionary studies.

The importance of Risk of Bias assessment

Here, we adopt the common view that RoB pertains to the internal validity of the study (Wang et al. 2022, but see Pescott et al. 2023). Internal validity refers to the extent to which a study can establish a causal relationship between the independent and dependent variables, free from confounding factors (Frampton et al. 2022). The key focus is on whether departures from the ideal causal study design could equally explain the finding. If it is plausible that they could, to varying degrees, then less trust may be placed in the calculated causal estimate. More broadly, the within-study RoB assessment can be a component of a critical appraisal, which can also include other quality aspects at the level of the primary study (e.g., external validity, reporting quality/transparency, methodological quality unrelated to bias; Juni et al. 2001; Sedgwick 2014; Guyatt et al. 2011), or biases at the level of the overall evidence synthesis (e.g., reviewer/researcher bias, biases in the evidence poll). We note that the quality of reporting in the primary studies will influence the opportunity to assess internal and external validity. Low reporting quality is the subject of existing guidelines in ecology and evolutionary biology (e.g., TOP guidelines, Parker et al. 2016a, 2016b). Unfortunately, the literature is frequently unclear on the definitions of the above-mentioned terms (risk of bias, critical appraisal, external

validity, biases in the evidence pool), which likely contributes to the confusion about their application and intent (cf. Wang et al. 2022). Box 1 lists the most common terms related to quality and bias in both primary studies and the overall evidence poll included in the evidence synthesis.

Threats to internal validity may result from biases in data collection, analysis, and results interpretation. Biases can vary in magnitude, and thus, their effect on the validity of a study's conclusions will also vary. However, RoB assessment usually aims to evaluate how susceptible primary studies are to biases rather than to quantify those biases, although quantitative bias assessments are well-developed in some disciplines (e.g., Lash et al. 2009). RoB assessment in systematic reviews and meta-analyses involves evaluating whether threats to internal validity are present in each included primary study, so that the findings of the primary studies and the systematic review or meta-analysis can then be interpreted in light of those potential biases.

RoB tools have been developed to facilitate and standardise RoB assessment, extensively so in medicine and especially for clinical trials. For example, out of 1172 systematic reviews that included randomised clinical trials and were published between 2023 and 2024, 1140 (97.3%) reported assessing the quality or risk of bias of the primary studies (Sandoval-Lentisco et al. 2024). Relatively close to ecology, the Center for Environmental Evidence has been developing a CEECAT tool (Konno et al. 2021). However, tools from other fields appear to be rarely used in ecology. A recent analysis of 680 systematic reviews in the field found that only 4% included any form of critical appraisal (Stanhope and Weinstein, 2023).

Box 1: Different biases and quality elements of the overall evidence synthesis, and of primary studies that can be included in the evidence synthesis.

Primary study (adapted from Frampton et al. 2022 and Stanhope & Weinstein 2023):

Internal validity - the extent to which a study can establish a causal relationship between the independent and dependent variables, free from confounding factors (Frampton et al. 2022). It will depend on the extent of any trend or systematic deviation in data collection, data analysis, and data interpretation. Internal validity is assessed as the Risk of Bias.

Precision (random error) - an error that arises from the inaccuracy of estimation that is inherent to probabilistic treatment assignment in research studies. Such an error is always present in estimates, especially when sample sizes are small. Sometimes, precision and internal validity together are referred to as "Methodological quality".

Reporting quality - the completeness of the reporting of methodological approaches, study results, and other elements such as research data. High reporting quality should allow for a complete reproduction of the study (i.e., all methodological aspects of data collection and analysis are fully described so they can be repeated). Results should be reported in full (e.g., mean value, error measure (confidence measure), and sample size).

Evidence synthesis:

Researcher (reviewer) bias - any bias that might influence the decisions of the researchers conducting the evidence synthesis. Researcher bias can affect all aspects of the evidence synthesis process, such as search strategy, exclusion/inclusion decisions, or the analytic approach. For instance, if a researcher conducting screening has a strong opinion on the direction of the effect, they might be more likely to exclude studies that do not support such an opinion. See, for example, Burghardt et al. (2012), Bishop (2017), and Ciria et al. (2023).

Evidence poll bias (literature bias) - a set of biases present in the overall poll of primary studies included in the evidence synthesis. These include publication bias, language bias, taxonomic biases (e.g., studies are predominantly conducted on a certain species), geographic biases, and similar biases. See, for example, Nakagawa et al. (2022), Konno et al. (2022), and Nunez & Amano (2021).

Risk of Bias - at the level of evidence synthesis, risk of bias (RoB) refers to the presence of biases within each primary study (i.e., the internal validity) included in the evidence synthesis. See Frampton et al. (2022).

Methodological quality of evidence synthesis - refers to the appropriateness of the search strategy, screening process, data extraction, and analysis (including assessment of the RoB and evidence poll bias). See Haddaway & Verhoeven (2015), Foo et al. 2021.

Reporting quality of evidence synthesis - refers to the completeness of reporting of all the methodological steps (e.g., search strings, search dates, databases used for search, inclusion and

exclusion criteria, number of researchers doing screening), results, and collected data used for the analysis. See more in Haddaway & Macura (2018), O'Dea et al. 2021.

Increasing the use of Risk of Bias assessment

Given their importance for the reliability of systematic reviews and meta-analyses, the question is "Why aren't RoB assessments commonly applied in ecology and evolution?". To start answering this question, we surveyed 232 ecologists and evolutionary biologists who have published systematic reviews and meta-analyses, and reviewed the guidelines of 275 journals that publish ecology and evolutionary biology research. Details can be found in the Methods section.

Based on the findings of the survey, previous research, and our own experience in the field, we recommend three main sets of actions to increase the use of RoB assessment in ecology and evolutionary biology.

1) Raising awareness and providing guidelines

An important barrier to RoB assessment in ecology and evolutionary biology might be a lack of awareness. While around half of the respondents of our survey (116 of 232, Fig. 1) had heard of the concept of RoB, only 28 correctly interpreted it (Fig. 1). Similarly, 66 respondents claimed to be aware of some RoB tools; however, more than half (37 of 66) confused them with publication bias tools. Indeed, while 53 participants answered that they had conducted a RoB assessment, the majority (34 of 53) had performed some form of publication bias test, and only 10 specifically mentioned an existing RoB tool they had used, or one that they had developed themselves (Fig 2.).

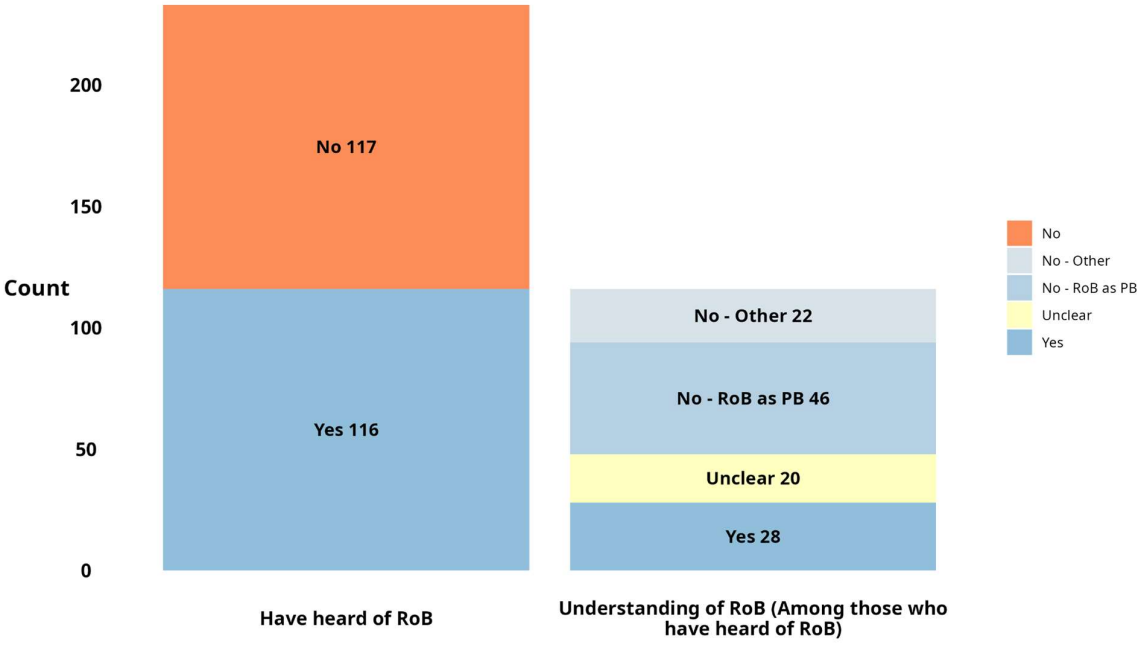

Figure 1. Survey results on RoB awareness among ecologists and evolutionary biologists with experience in evidence synthesis. Left-hand bar: respondents who claimed to have heard of RoB ("Yes"), and have not heard of RoB ("No"). Right-hand bar: respondents who had a correct understanding of RoB ("Yes"), those who confused RoB with publication bias ("No - RoB as PB"), had some other incorrect interpretation ("No - Other"), and those for whom we could not judge their level of understanding based on the responses ("Unclear"). The figure can be reproduced using data and code from Culina et al. 2025.

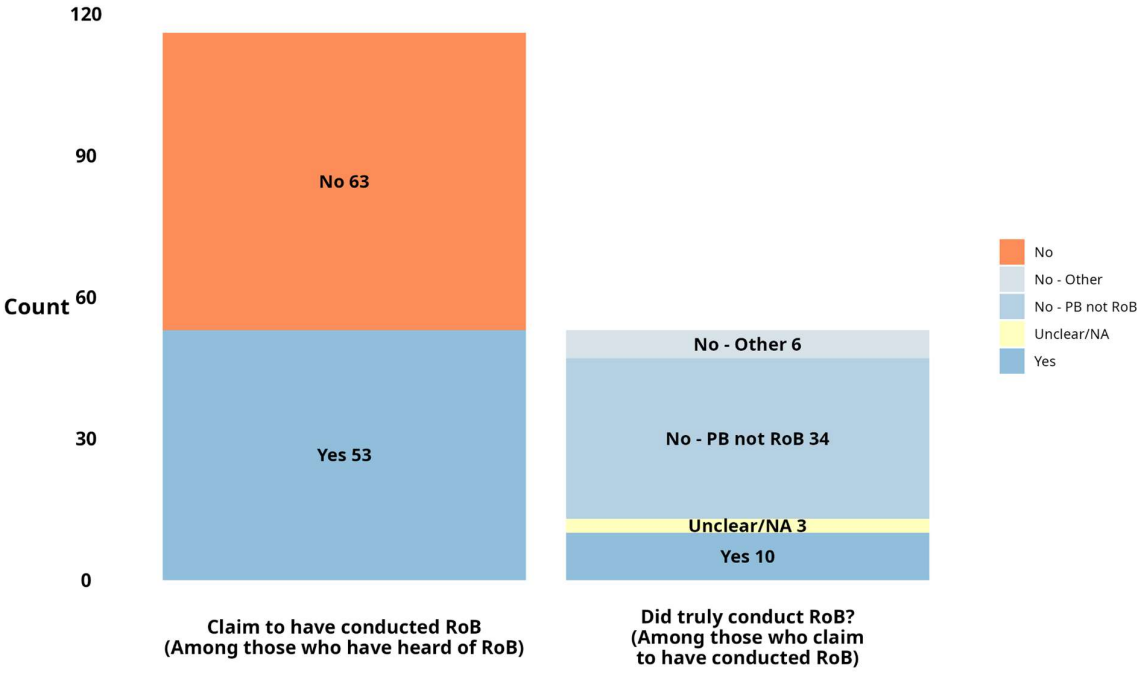

Figure 2. Survey results on experience in conducting RoB assessment among ecologists and evolutionary biologists who were aware of RoB. Left-hand bar: the respondents who claimed to have conducted a RoB assessment ("Yes") and have not conducted it ("No"). Right-hand bar: the respondents who truly conducted RoB assessment ("Yes"), those who conducted a publication bias test ("No - PB not RoB''), conducted some other test/assessment ("No - Other"), and those with responses from which we could not judge what they have conducted ("Unclear"). The figure can be reproduced using data and code from Culina et al. 2025.

In addition, between 19.04.2025 and 12.05.2025, we checked the websites of 275 journals that publish ecology and evolutionary biology research (see Methods for details). Out of 209 journals that either explicitly solicit evidence synthesis (58 journals) or solicit 'Reviews' (151 journals), only five mentioned standards of conducting evidence synthesis, which include RoB or a related (e.g. critical appraisal) assessment (Fig. 3). These journals either then link to guidelines on critical appraisal assessment directly (e.g., CEE guidelines) or to reporting guidelines (e.g., PRISMA) that, within them, mention the need for the RoB assessment. An

additional 45 journals mention reporting guidelines for evidence synthesis (e.g., PRISMA), but in reference to the evidence synthesis reporting rather than to assessing RoB.

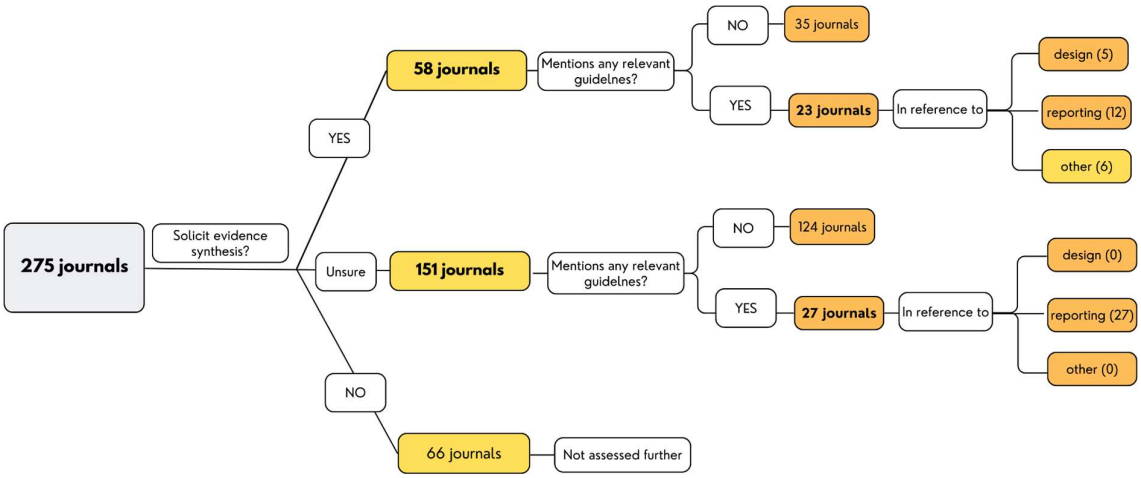

Figure 3. Summary results on the mention of RoB or related assessment by 275 journals that publish ecological and evolutionary biology research. We first assessed if journals solicit evidence synthesis ("YES") or just vaguely refer to "Reviews" ("Unsure"). We then checked if they mention any guidelines, either that directly address RoB assessment or related (i.e., critical appraisal), or other guidelines (e.g., reporting guidelines) that within themselves mention RoB or related. Bolded text represents journals that were assessed in the next steps.

Based on the findings above, we call for the evidence synthesis community to increase awareness of the importance of RoB assessment in ecology and evolutionary biology. First, RoB assessment should be included in evidence synthesis teaching and workshop materials. Second, journals and funders should explicitly mention RoB assessment in their guidelines and requirements for evidence synthesis, and raise awareness by publishing editorials, perspectives, and other types of articles. Related to this, we noticed that most of the journals (151) just mention accepting "Reviews", without further specifying whether these include systematic reviews or only narrative reviews. Being explicit on whether a journal accepts systematic reviews, meta-analyses, and other types of evidence synthesis, together with clear guidelines on conducting and reporting these, would be an important step in ensuring high-quality evidence synthesis.

2) Facilitating RoB assessment

The RoB tools developed in other fields seem not to be easily applicable to most systematic reviews and meta-analyses in ecology and evolutionary biology. In addition, they are time-consuming to apply, given how much detail they require, which likely contributes to their limited implementation in ecology and evolutionary biology. For example, Crocker et al. (2023) applied Version 2 of the Cochrane RoB tool for randomised controlled trials to a systematic review, which took an average of 358 minutes per study. Since the average meta-analysis in ecology and evolutionary biology includes about 40 studies (median = 41 studies in plant ecology, Koricheva and Gurevitch 2014; 44 in evolutionary ecology, Pollo et al. 2024), that could constitute around 240 hours of additional work per meta-analysis or thirty 8-hour days. While large centres dedicated to evidence synthesis (e.g., Cochrane, Campbell) also have large teams who only conduct evidence synthesis, and who are trained in all aspects of its methodology (including RoB assessment), most of the ecological and evolutionary synthesis is done by less specialised, small teams. Small team size combined with non-tailored yet detailed

tools might further lead to low use of RoB assessments in ecology and evolutionary biology. Out of 10 respondents who have conducted RoB assessment in their meta-analysis, 8 have judged it to be relatively difficult (5 or more on a scale of 1 to 10, where 1 is 'very easy' and 10 is 'very difficult'). Despite this, 7 respondents said they were likely to conduct a RoB assessment again.

Creating tools that would fit the requirements of a wide range of research studies in ecology and evolutionary biology is challenging. The first step would be to gather a group of evidence synthesis experts in ecology and evolutionary biology to identify all possible study designs and to propose the relevant RoB items to be assessed for these study designs. In Box 2, we provide the main steps in creating a robust RoB tool. In section 'Recommendations for a basic RoB evaluation', we propose four main general questions researchers might want to consider as a part of a RoB assessment, or when designing a primary study.

3) Making RoB possible - reporting standards

A big barrier when assessing RoB is low reporting quality in primary studies, which makes some of the RoB items difficult or impossible to judge. For example, primary studies in ecology and evolutionary biology rarely explicitly mention whether researchers were blinded to data collection or analysis (e.g., Kim et al. 2021; Mentesana et al. 2025). Similarly, under-reporting important details of methods, such as sample sizes, is common in ecology and evolution (Parker 2016b). Such poor reporting not only makes RoB assessment difficult or impossible but can also lead to exclusion of primary studies from meta-analyses (e.g., Moran et al. 2021; Santostefano et al. 2025), which has led to several contributions explaining how to report primary studies for inclusion in evidence synthesis (e.g., Gerstner et al. 2017; Henessy et al. 2021). Usually, when a RoB item cannot be scored due to insufficient reporting (i.e., rated as 'unclear') the study is considered to have failed to mitigate that specific RoB.

Journals play a crucial role in establishing detailed reporting guidelines for both methods and results of published (primary) studies. For instance, to improve the transparency of reporting and the reproducibility of published results, authors of research papers in life sciences, behavioural & social sciences and ecology, evolution & environmental sciences submitting to *Nature* portfolio journals must provide a completed reporting summary that will be made available to editors and reviewers during manuscript assessment and published with all accepted manuscripts. Other examples include Ecology letters that in 2012 introduced reporting standards in experimental studies (Hillebrand and Gurevitch 2013), and *Functional Ecology* that has recently introduced the requirement of a 'replication statement' to be included in the Materials and Methods section.

Box 2: How to create Risk of Bias tools

Given the difficulty of reaching universal agreement on definitions of study quality and related attributes (Higgins et al. 2011; Wang et al. 2022), RoB tools are typically developed through consensus-building exercises. A typical scenario is for project leads to draw up long lists of potential RoB domains. This is followed by larger Delphi-technique-type processes (Mukherjee et al. 2015), in which collaborators review evidence, discuss uncertainties, and ultimately seek consensus on RoB source inclusion, definitions, supporting descriptions, guidance, tool "signalling questions" and structure (e.g., Higgins et al. 2011; Boyd et al. 2022). Signalling questions have been described as "reasonably factual questions with yes/no answers that inform RoB judgments" (Sterne et al. 2019). Frampton et al. (2022) described the process of developing RoB tools as "constantly evolving", pointing to the "target study" paradigm for assessing the RoB in non-randomised studies, where a focal trial is compared to a Platonic conception of the perfect randomised target study for answering the question of interest (the perfect study imagined may not even be possible to conduct in the real world; Sterne et al. 2016).

Once the tool has been designed, it has to be evaluated to enhance consistency in its interpretation and application, and to make it user-friendly. Researchers have progressively evaluated RoB tools by measuring their inter-rater reliability, and have identified tool size and complexity, clarity of user guidance, and the necessity of individual judgement as key sources of variability (Hartling et al. 2009, 2013; Minozzi et al. 2019). Evaluations have also measured the time costs of applying different tools

and the time taken for reviewer teams to resolve conflicting assessments and reach consensus (Jeyaraman et al. 2020).

Tools are commonly upgraded, drawing on user feedback and other considerations. For example, the update to the Cochrane RoB tool for randomised controlled trials, RoB2 (Sterne et al. 2019), involved reviewing how the original tool of Higgins et al. (2011) had been used, an empirical systematic review and meta-analysis of evidence synthesis efforts in epidemiology focused on associations between bias domains and effect sizes (Page et al. 2016) and a separate systematic review of the "theoretical and conceptual literature on types of bias in epidemiology" (Higgins et al. 2019); other sources of evidence relating to the types and frequencies of biasing domains in medical research were also used (e.g., Page & Higgins, 2016).

Frampton et al. (2022) emphasised that new tool creation is not always necessary: fundamental study design types are limited across disciplines, and therefore, bias domains tend to be repeated as well. Frampton et al. (2022) recommend using FEAT (Focused, Extensive, Applied, Transparent) principles in this decision-making:

(i) Does a tool Focus clearly on RoB as opposed to other broad quality constructs such as reporting guideline compliance?;

(ii) Is the tool Extensive in its coverage? That is, does it cover all possible sources of RoB for the study type?;

(iii) Can the RoB tool outcomes be Applied in a useful way? That is, can outcomes be clearly and logically incorporated into evidence synthesis reporting?; and

(iv) Are the RoB judgements required by the Tool explicit and unambiguous (i.e., Transparent)?

Recommendations for a basic Risk of Bias assessment

Before specific RoB tools are developed for ecology and evolutionary biology, researchers can start with a basic RoB assessment of primary studies (or effect sizes within the primary studies). Here, we suggest a list of four basic questions for assessing RoB. We based these on existing RoB assessment tools (CEECAT, Konno et al. 2021; ROBINS-I, Sterne et al. 2016) and the Catalogue of Bias (https://catalogofbias.org/). We suggest each question (for each study, or each effect size within the study) should be answered as: Low RoB, High RoB, Unclear RoB,

or Not Applicable (i.e., the type of bias is not relevant for the type of study assessed). The "Unclear" category is used when RoB cannot be confidently assessed, usually because a study did not report its methodology completely. The category "Unclear" then puts less trust into the study than 'Low RoB', and is often considered as 'High RoB'. As is the case for screening and data extraction (Foo et al. 2021), the scoring of each study should ideally be done by more than one person after a pilot stage that guarantees inter-rater reliability, and any disagreements should be reported, discussed and resolved.

After primary studies are scored based on these four questions, the analysis (usually meta-analysis) could include only studies with a low RoB (low overall score), or test if the inclusion of high RoB studies impacts the direction, magnitude or precision of effects in a sensitivity analysis. This is a standard methodological expectation of Cochrane and Campbell reviews. Another approach is model-based weighting using RoB scores; however, such approaches may be questionable as they are unlikely to directly map to actual study-specific drivers of bias (see Pescott and Stewart 2022).

The RoB questions we suggest are:

1) Was randomisation applied?

In experiments, randomisation involves assigning experimental units (e.g., individuals, nests, plots) randomly to each treatment. Randomisation minimises differences between the experimental units in the control and treatment groups (but see Senn, 2013). It is the foundation of experimental causal inference, and randomisation-based study designs reduce the likelihood of biased treatment effect estimators. In observational studies, units are not assigned to exposure by the investigator; instead, causal inference depends on design or statistical strategies (e.g., instrumental variables, propensity-score methods, regression discontinuity) to approximate random assignment. The method of sampling (whether random or not) from the

population is normally considered separately as a factor influencing generalisability, not as a part of the randomisation-for-causal-inference step (although for descriptive inference it can be argued that non-probabilistic sampling is the main RoB; Boyd et al. 2022; Pescott et al. 2023). Therefore, whether or not a study has adequately randomised exposure (in experimental studies), or approximated random exposure via design or analysis methods (in observational studies), provides important information regarding the RoB of the estimates it provides. A general overview of the importance of randomisation in causal inference can be found in Greenland (1990).

In many ecological and evolutionary studies, complete randomisation is impossible. For example, land areas might not be equally available or accessible to sample from. In mark-recapture studies, some animals might be less detectable, which might be linked to a trait of interest (e.g., Culina et al. 2013). In this case, analytical approaches might partly account for the issue, although inferential certainty might be weakened (e.g., yielding lower precision). We acknowledge that unequal detectability in a mark–recapture study does not reflect a failure to randomise per se but rather differential selection or measurement bias at the sampling stage. Formal RoB tools would typically assess such issues under 'selection bias' rather than as a breakdown of random allocation. However, for simplicity in this introductory paper, we lump these considerations together.

2) Was confounding considered?

This question is relevant if the previous question was answered with "No" (i.e. "High risk of bias" or "Unclear". Biases can arise due to an uncontrolled (or inappropriately controlled) variable (confounder) that influences both the intervention/exposure and the outcome (e.g., see Marshall 2024); including studies where no or an inappropriate control group is used (e.g., Engqvist and Reinhold 2016). However, it can be difficult to control for potential confounders

in observational studies, often because many confounders remain unknown. Furthermore, compared to meta-analyses of clinical trials, where studies included are often more standardised, and almost always conducted on the same species, studies included in meta-analyses in ecology and evolutionary biology are commonly very diverse (e.g., different species, different habitats), which can lead to confounders likely differing between the studies. If the studies did control for the confounders in the analysis, but confounders are different between the studies, such 'diverse' effect sizes might not be possible to include in the meta-analysis. Thus, combining results from ecological and evolutionary biology studies that use either inappropriate confounders or different sets of confunders among them may impact the overall reliability of the meta-analytic estimates (Velentine and Thomoson 2012). Confounding bias can also occur when meta-analyses include primary study estimates for the effects of variables originally modelled as controls (Westreich and Greenland, 2013). Petersen et al. (2022) and Konno et al. (2021) provide tools/details on assessing confounding bias in systematic reviews of observational studies, which might be a useful reference to read for those who would like to reduce or assess confounder bias.

3) Were the methods for measuring the outcome data the same across the groups/individuals?

Biases can arise from systematic differences between the true and the measured value, where the differences are group-specific (caused by the observer or measuring mode). For example, two groups are compared based on a certain trait. The measurement in group A is always taken by observer A, and the measurement in group B is always taken by observer B. Here, each observer might measure in a slightly different way, especially if observers have not received common training in how to obtain a measurement.

4) Was the observer or analyst blind to the allocation of subjects?

In ecology and evolutionary biology, blinding is rarely applied (less than 20% of 981 studies, Purgar et al. 2022; 6.3% of 960 studies, Burghardt et al. 2012). Yet, evidence gathered so far suggests that unblinded data collection leads to inflated effect sizes compared to blinded data collection (Holtmann et al. 2015, Kardish et al. 2015).

Blinding oneself during data collection is often not possible or extremely impractical in ecology and evolutionary biology, especially in field studies. In this case, researchers can usually blind themselves during data analysis. Further, non-blinding might not be a problem if a study uses objective tools to judge outcomes (e.g., a quantitative outcome derived from the application of some measuring device). However, even if blinding is difficult to implement in any given study and/or quantitative measurements are made, the resulting RoB remains non-zero: unconscious expectations can still subtly influence data processing or interpretation even when a device generates numeric outputs (MacCoun and Perlmutter, 2015).

In all, we have provided four RoB questions that most researchers should be able to answer when judging the RoB of primary studies. However, many existing, more detailed tools could be consulted, adjusted or applied when conducting a RoB assessment in ecology and evolutionary biology. Some of these include the CEECAT tool (Konno et al. 2021), ROBITT (Boyd et al. 2022), SYRCLE (Hooijmans et al. 2014), and Robins-I (Sterne et al. 2016). Use of non-standard tools utilising review specific criteria is also an approach that has been used. As long as questions are asked and answered transparently, readers can judge the appropriateness for themselves. However, we caution against application of scoring scales or generic checklists that conflate reporting or precision or other methodological components with assessment of risk of bias or internal validity.

Conclusions

Risk of Bias (RoB) assessment is essential for the transparency and quality of evidence synthesis results, and enables explorations of effect heterogeneity and reliability. RoB assessment should therefore be an essential part of any systematic review and meta-analysis, and its use should be extended to anyone planning a primary study or critically reading research articles. To increase its application in ecology and evolutionary biology, we advocate for elevating the importance of RoB assessment through different channels (training, guidelines, and policies). It is time for the evidence synthesis community within ecology and evolutionary biology to come together and start developing a RoB tool(s) more suitable for these fields. To start with, we suggest four questions that can be used to cover the main aspects of RoB assessment.

Journals should provide explicit statements on whether they accept systematic reviews, meta-analyses, and other evidence synthesis contributions. These statements should be explicitly made in the "Aims and Scope" or under "Article types". Those who accept evidence synthesis should provide clear guidance on the highest quality standards by explicitly linking to existing best practices on conducting and reporting evidence synthesis. Journals should further aim to improve the reporting quality in primary studies (as previously recommended, e.g., Ihle et al. 2017; Gerstner et al. 2017; Hennessy et al. 2021), which would enable easier assessment of RoB. This means both requiring authors to follow reporting guidelines or providing reporting checklists, and ensuring these are followed and provided. Finally, we note that many RoB elements will be very difficult to fully address in ecological and evolutionary studies, especially in observational ones. This does not mean such studies should be discouraged, but such limitations need to be considered when making inferences.

Methods

1) Survey on the RoB awareness and use

The survey was approved by the Ethics Committee of the Ruđer Bošković Institute, Zagreb, Croatia, ref. ZV/3218/1-2023. The survey was aimed at ecologists and evolutionary biologists who have published at least one meta-analysis. It included non-identifying questions on familiarity with the concept of the Risk of Bias, awareness and use of RoB assessment, and questions on familiarity with meta-analysis, field of research, and career stage (survey questions can be found in the Supplementary Materials).

The survey was created in Google Forms and sent on the 11th September 2023 to the emails of corresponding authors of meta-analyses in ecology and evolution, via mailing lists (NC3 Collaborative Research Centre; Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology; German Zoological Society; Sociedad Española de Etología y Ecología Evolutiv;, EuropeList@conbio.org), slack channels (Big Team Science Conference Slack Channel, German Reproducibility Network Event Slack Channel, ESMARConf Slack Channel), Twitter posts, and the SORTEE newsletter. The survey was open to responses until the 15th October 2023. The invitation email text is available in the Supplementary Materials.

To determine the corresponding authors of meta-analyses, AST (author) searched for meta-analyses published in 300 journals via Web of Science (databases: SCI-EXPANDED, SSCI, AHCI, ESCI) on the 25th April 2023 (search string, which also provides codes for included journals, can be found in the Supplementary materials and from Culina et al. 2025). This search retrieved 3,289 results (potential meta-analyses, BibTeX list of these available from Culina et al. 2025) published between 1945 and 2023. AST then extracted a list of all corresponding author email addresses from these articles using packages revtools 0.4.1 (Westgate 2019a, 2019 b) and stringr 1.5.0 (Wickham 2023) in R 4.2.3 (R Core Team, 2021). Code is available from

Culina et al. 2025. This resulted in 3,346 email addresses; however, 789 emails (~24%) bounced back.

The survey received 232 valid responses (i.e., 9.1% response rate). Because some of the responses were prone to subjective judgment on what the answer exactly meant, four assessors (AC, AST, ROD, and MG; authors) went through the answers of 188 respondents who had answered 'YES' to the question 'Prior to receiving this survey invite, were you familiar with the concept of Risk of Bias (RoB)?' or who had answered 'NO' but their remaining answers indicated otherwise. Each assessor provided an answer 'YES', 'NO', 'Unsure', or 'NA' to the following six questions:

1) Has the respondent heard of RoB?
2) Does the respondent have a correct interpretation of RoB?
3) Does the respondent claim to have conducted a RoB assessment?
4) Have they truly conducted RoB assessment?
5) Respondent thinks RoB is publication bias
6) Respondent has conducted publication bias, rather than RoB assessment

We then compared the interpretations of the responses among all four assessors. When three or four assessors had the same interpretation, we chose the most common answer as the final one. When there was disagreement (usually a 50:50 split; this happened for 26 out of 708 entries), we discussed the interpretations and agreed on whether the final answer should be 'YES', 'NO', or 'NA' (i.e., the question is irrelevant given the previous answers, or we could not agree on the interpretation, or we agreed that the answer was too vague to interpret). The original data table with separate scores for each reviewer and the table with the final (combined) scores are available from Culina et al. 2025.

2) Journals and RoB assessment

Between 19th April 2025 and 12th May 2025, AC, OP, and ROD (authors) checked the websites of 275 journals that publish ecology and evolutionary biology research. The list of journals was taken from Ivimey-Cook et al. (2025), and is also available in the relevant data table from Culina et al. 2025. We checked each journal's Aims & Scope, Author instructions, and Editorial policy sections to search for whether a journal accepts evidence synthesis, and whether it mentions Risk of Bias or any related concepts for authors of evidence synthesis articles.

We first piloted our data extraction on 10 journals to adjust the data extraction questions and align responses. We used Google Forms for data extraction. The exact extraction questions can be found in the Supplementary Materials. The main questions included:

1) Does the journal explicitly solicit some form of evidence synthesis?

2) Does the journal specifically mention RoB or related assessment of primary literature in guidelines to authors of evidence synthesis, or is RoB/related assessment specifically mentioned in linked other guidelines (e.g. journal states something like 'follow PRISMA guidelines when reporting MA' but no further detail on RoB or related is mentioned)?

3) What concept related to RoB (including RoB itself) does the journal or a linked guideline exactly mention?

4) If the journal links or refers to external guidelines that mention RoB or related assessment, what are these guidelines?

5) What specific RoB or related tool/checklist is mentioned in journal guidelines, or in linked guidelines?

6) What is the strength of the journal's policy on the use of RoB or related assessment?

The following decisions were made in light of the pilot extraction: First, we followed this definition of evidence synthesis ‘‘Evidence syntheses are conducted in an unbiased, reproducible way to provide evidence for practice and policy-making, as well as to identify gaps in the research. Evidence syntheses may also include a meta-analysis, a more quantitative process of synthesising and visualising data retrieved from various studies’ (https://guides.library.cornell.edu/evidence-synthesis). Second, journals that explicitly solicit narrative reviews and similar (e.g. the Annual Review of Ecology, Evolution and Systematics solicits ‘essay reviews’ but not systematic or quantitative reviews) were scored as ‘NO’ for question (1) above, whereas journals that do not explicitly solicit evidence synthesis, but something more general (e.g. review articles, reviews and comprehensive synthesis, reviews) were scored as ‘Unsure’ for the same question.

We divided journals across reviewers, and one reviewer checked each journal. The data table containing data reviewer initials and their scores can be found in Culina et al. 2025.

Funding

This work was supported by the Croatian Science Foundation under the project number HRZZ- IP-2022-10-2872 to A. Čulina.

Conflict of interest

Authors declare no conflict of interest.

References

Bishop, D. V. M. (2017). *Standing on the shoulders of giants: Why your literature review should be systematic*. https://osf.io/uk92g.

Boyd, R.J. *et al*.(2022). "ROBITT : A Tool for Assessing the Risk-of-bias in Studies of Temporal Trends in Ecology." *Methods in Ecology and Evolution* 13 (7): 1497–1507. https://doi.org/10.1111/2041-210X.13857.

Burghardt, G.M. *et al*. (2012). "Perspectives – Minimizing Observer Bias in Behavioral Studies: A Review and Recommendations." *Ethology* 118 (6): 511–17. https://doi.org/10.1111/j.1439-0310.2012.02040.x.

Ciria, Luis F. *et al.* (2023). "An Umbrella Review of Randomized Control Trials on the Effects of Physical Exercise on Cognition." *Nature Human Behaviour* 7 (6): 928–41. https://doi.org/10.1038/s41562-023-01554-4.

Cooke, S.J. *et al.* (2017). "How Experimental Biology and Ecology Can Support Evidence-Based Decision-Making in Conservation: Avoiding Pitfalls and Enabling Application." *Conservation Physiology* 5(1) https://doi.org/10.1093/conphys/cox043.

Crocker, T.F. *et al.* (2023). "Risk-of-Bias Assessment Using Cochrane's Revised Tool for Randomized Trials (RoB 2) Was Useful but Challenging and Resource-Intensive: Observations from a Systematic Review." *Journal of Clinical Epidemiology* 161: 39–45. https://doi.org/10.1016/j.jclinepi.2023.06.015.

Culina, A., Lachish, S., Pradel, R., Choquet, R. and Sheldon B.C. (2013) "A Multievent Approach to Estimating Pair Fidelity and Heterogeneity in State Transitions." *Ecology and Evolution* 3 (13): 4326–38. https://doi.org/10.1002/ece3.729.

Culina, A., Sánchez-Tójar, A., Pescott, O., O'Dea, R., & Grainger, M. (2025). Datasets and codes for 'Elevating the importance of Risk of Bias assessment for ecology and evolution' [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15630232

Else, H.. "'Systematic Reviews' That Aim to Extract Broad Conclusions from Many Studies Are in Peril," November 28, 2024. https://doi.org/10.1126/science.zpnivp6.

Engqvist, L. and Reinhold, K. (2016). Adaptive trans-generational phenotypic plasticity and the lack of an experimental control in reciprocal match/mismatch experiments. *Methods Ecol Evol* 7: 1482-1488. https://doi.org/10.1111/2041-210X.12618

Foo, Y. Z., O'Dea, R.E., Koricheva, J., Nakagawa, S., and Lagisz, M. (2021). "A Practical Guide to Question Formation, Systematic Searching and Study Screening for Literature Reviews in Ecology and Evolution." *Methods in Ecology and Evolution* 12 (9): 1705–20. https://doi.org/10.1111/2041-210X.13654.

Frampton, G. *et al.* (2022). "Principles and Framework for Assessing the Risk of Bias for Studies Included in Comparative Quantitative Environmental Systematic Reviews." *Environmental Evidence* 11 (1): 12. https://doi.org/10.1186/s13750-022-00264-0.

Gerstner, K. *et al.* (2017). "Will Your Paper Be Used in a Meta-analysis? Make the Reach of Your Research Broader and Longer Lasting." Edited by David Warton. *Methods in Ecology and Evolution* 8 (6): 777–84. https://doi.org/10.1111/2041-210X.12758.

Grainger, M.J., Bolam, F.C, Stewart, G.B., and Nilsen, E.B. (2020). "Evidence Synthesis for Tackling Research Waste." *Nature Ecology & Evolution* 4 (4): 495–97. https://doi.org/10.1038/s41559-020-1141-6.

Greenland, S. (1990) "Randomization, Statistics, and Causal Inference." *Epidemiology* 1, no. 6: 421–29.

Guyatt, G.H., *et al*. (2011). "GRADE Guidelines: 8. Rating the Quality of Evidence—Indirectness." *Journal of Clinical Epidemiology* 64(12): 1303–10. https://doi.org/10.1016/j.jclinepi.2011.04.014.

Haddaway, N.R., Macura, b., Whaley, p. and Pullin, A.S. (2018). "ROSES RepOrting Standards for Systematic Evidence Syntheses: Pro Forma, Flow-Diagram and Descriptive Summary of the Plan and Conduct of Environmental Systematic Reviews and Systematic Maps." *Environmental Evidence* 7 (1): 7. https://doi.org/10.1186/s13750-018-0121-7.

Haddaway, N. R., and Verhoeven, J.T.A. (2015). "Poor Methodological Detail Precludes Experimental Repeatability and Hampers Synthesis in Ecology." *Ecology and Evolution* 5(19): 4451–54. https://doi.org/10.1002/ece3.1722.

Haddaway, N.R., and Macura, B. (2018). "The Role of Reporting Standards in Producing Robust Literature Reviews." *Nature Climate Change* 8(6): 444–47. https://doi.org/10.1038/s41558-018-0180-3.

Hartling, L. *et al.* (2009). "Risk of Bias versus Quality Assessment of Randomised Controlled Trials: Cross Sectional Study." *BMJ* 339: b4012–b4012. https://doi.org/10.1136/bmj.b4012.

Hartling, L. *et al.* (2013). "Testing the Risk of Bias Tool Showed Low Reliability between Individual Reviewers and across Consensus Assessments of Reviewer Pairs." *Journal of Clinical Epidemiology* 66(9): 973–81. https://doi.org/10.1016/j.jclinepi.2012.07.005.

Hennessy, E.A. et al. (2022). Ensuring Prevention Science Research is Synthesis-Ready for Immediate and Lasting Scientific Impact. *Prev Sci* 23: 809–820. https://doi.org/10.1007/s11121-021-01279-8

Higgins, J. P. T. *et al.* (2011). "The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials." *BMJ* 343: d5928–d5928. https://doi.org/10.1136/bmj.d5928.

Hillebrand, H. and Gurevitch, J. (2013). Reporting standards in experimental studies. *Ecology Letters* 16: 1419–1420

Holman, L., Head, M.L., Lanfear, R., and Jennions, M.D. (2015). "Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording." *PLOS Biology* 13(7): e1002190. https://doi.org/10.1371/journal.pbio.1002190.

Hooijmans, C.R. *et al.* (2014). SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 14(43). https://doi.org/10.1186/1471-2288-14-43

Ihle, M., Winney, I.S., Krystalli, A. and Croucher, M. (2017). "Striving for Transparent and Credible Research: Practical Guidelines for Behavioral Ecologists." *Behavioral Ecology* 28(2): 348–54. https://doi.org/10.1093/beheco/arx003.

Jeyaraman, M.M. *et al.* (2020). "Methodologically Rigorous Risk of Bias Tools for Nonrandomized Studies Had Low Reliability and High Evaluator Burden." *Journal of Clinical Epidemiology* 128: 140–47. https://doi.org/10.1016/j.jclinepi.2020.09.033.

Juni, P. (2001). "Systematic Reviews in Health Care: Assessing the Quality of Controlled Clinical Trials." *BMJ* 323: 42–46. https://doi.org/10.1136/bmj.323.7303.42.

Kardish, M.R., Mueller, U.G., Amador-Vargas, S., Dietrich, E.I., Ma, R., Barrett, B., and Fang, C-C (2015). "Blind Trust in Unblinded Observation in Ecology, Evolution, and Behavior." *Frontiers in Ecology and Evolution* 3 https://doi.org/10.3389/fevo.2015.00051.

Konno, K, Livoreil, B, and Pullin, A.S. (2021). "Collaboration for Environmental Evidence Critical Appraisal Tool Version 0.3." Collaboration for Environmental Evidence. https://environmentalevidence.org/cee-critical-appraisal-tool/.

Konno, K. *et al.* (2020). "Ignoring non-English-language Studies May Bias Ecological Meta-analyses." *Ecology and Evolution* 10(13): 6373–84. https://doi.org/10.1002/ece3.6368.

Koricheva, J. and Gurevitch, J. (2014)."Uses and misuses of meta-analysis in plant ecology." *J Ecol* 102: 828-844. https://doi.org/10.1111/1365-2745.12224

Lash, T.L., Fox, M.P., and Fink, A.K. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer Science & Business Media, 2009.

Lodi, S., Spacek Godoy B, Carlo Gonçalves Ortega J, Mauricio Bini L. (2021). "Quality of meta-analyses in freshwater ecology: A systematic review." *Freshw Biol.* 66: 803–814. https://doi.org/10.1111/fwb.13695

MacCoun, R. and Perlmutter, S.. (2015). "Blind analysis: Hide results to seek the truth." *Nature* 526(7572): 187-189.

Marshall, D.J. (2024). "Principles of Experimental Design for Ecology and Evolution." *Ecology Letters* 27(4): e14400. https://doi.org/10.1111/ele.14400.

Martin, L.J., Blossey, B. and Ellis, E. (2012). "Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations." *Frontiers in Ecology and the Environment* 10: 195-201. https://doi.org/10.1890/110154

Mentesana, L., Hau, M., D'Amelio, P.B., Adreani, N.M., and Sánchez-Tójar, A. (2025). "Do Egg Hormones Have Fitness Consequences in Wild Birds? A Systematic Review and Meta-Analysis." *Ecology Letters* 28(3): e70100. https://doi.org/10.1111/ele.70100.

David, M., Liberati, A., Tetzlaff, J., Altman, D.G., and The PRISMA Group (2009). "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLoS Medicine* 6(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Moran, N.P., Sánchez-Tójar, A., Schielzeth, H. and Reinhold, K. (2021). "Poor nutritional condition promotes high-risk behaviours: a systematic review and meta-analysis.". *Biol Rev* 96: 269-288. https://doi.org/10.1111/brv.12655

Mukherjee, N. et al. (2015). "The Delphi Technique in Ecology and Biological Conservation: Applications and Guidelines." Edited by Barbara Anderson. *Methods in Ecology and Evolution* 6(9): 1097–1109. https://doi.org/10.1111/2041-210X.12387.

Nakagawa, S. *et al.* (2020). "A New Ecosystem for Evidence Synthesis." *Nature Ecology & Evolution* 4 (4): 498–501. https://doi.org/10.1038/s41559-020-1153-2.

Minozzi, S., Cinquini, M., Gianola, S., Castellini, G., Gerardi, C. and Banzi, R. (2019). "Risk of Bias in Nonrandomized Studies of Interventions Showed Low Inter-Rater Reliability and Challenges in Its Application." *Journal of Clinical Epidemiology* 112: 28–35. https://doi.org/10.1016/j.jclinepi.2019.04.001.

Nakagawa, S. et al. (2022). "Methods for Testing Publication Bias in Ecological and Evolutionary Meta-analyses." *Methods in Ecology and Evolution* 13(1): 4–21. https://doi.org/10.1111/2041-210X.13724.

Nuñez, M.A., and Amano, T. (2021). "Monolingual Searches Can Limit and Bias Results in Global Literature Reviews." *Nature Ecology & Evolution* 5(3): 264–264. https://doi.org/10.1038/s41559-020-01369-w.

O'Dea, R.E. *et al.* (2021). "Preferred Reporting Items for Systematic Reviews and Meta-analyses in Ecology and Evolutionary Biology: A PRISMA Extension." *Biological Reviews* 96(5): 1695–1722. https://doi.org/10.1111/brv.12721.

Page, M.J. *et al.* (2021). "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *BMJ* 71. https://doi.org/10.1136/bmj.n71.

Page, M.J., and Higgins, J.P.T. (2016). "Rethinking the Assessment of Risk of Bias Due to Selective Reporting: A Cross-Sectional Study." *Systematic Reviews* 5(1): 108. https://doi.org/10.1186/s13643-016-0289-2.

Parker, T. H., Nakagawa, S., Gurevitch, J. and IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants. (2016) "Promoting Transparency in Evolutionary Biology and Ecology." *Ecology Letters* 19(7): 726–28. https://doi.org/10.1111/ele.12610.

Parker, T.H. *et al.* (2016). "Transparency in Ecology and Evolution: Real Problems, Real Solutions." *Trends in Ecology & Evolution* 31(9): 711–19. https://doi.org/10.1016/j.tree.2016.07.002.

Pescott, O.L., Boyd, R.B., Powney, G.D., and Stewart., G.B. (2023) "Towards a Unified Approach to Formal Risk of Bias Assessments for Causal and Descriptive Inference." arXiv . https://doi.org/10.48550/ARXIV.2308.11458.

Pescott, O.L., and Stewart, G.B. (2022). "Simulation-based Study Design Accuracy Weights Are Not Generalisable and Can Still Lead to Biased Meta-analytic Inference: Comments on Christie et al. (2019)." *Journal of Applied Ecology* 59(5): 1187–90. https://doi.org/10.1111/1365-2664.14153.

Petersen, J.M. et al. (2022). “The Confounder Matrix: A Tool to Assess Confounding Bias in Systematic Reviews of Observational Studies of Etiology.” *Research Synthesis Methods* 13(2): 242–54. https://doi.org/10.1002/jrsm.1544.

Purgar, M., Klanjscek, T., and Culina A. (2022). “Quantifying Research Waste in Ecology.” *Nature Ecology & Evolution* 6(9): 1390–97. https://doi.org/10.1038/s41559-022-01820-0.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sandoval-Lentisco, A., López-López, J.A., and Sánchez-Meca, J. (2024). “Frequency of Use of the Revised Cochrane Risk of Bias Tool (RoB 2) in Cochrane and non-Cochrane Systematic Reviews Published in 2023 and 2024.” *Research Synthesis Methods* 15(6): 1244–45. https://doi.org/10.1002/jrsm.1755.

Santostefano, F., Moiron, M., Sánchez-Tójar, A., Fisher, D.N. (2025). “Indirect genetic effects increase the heritable variation available to selection and are largest for behaviors: a meta-analysis“. *Evolution Letters* 9(1), 89–104, https://doi.org/10.1093/evlett/qrae051

Sedgwick, P. (2014). “Randomised Controlled Trials: Internal versus External Validity.” *BMJ* 348: 1742–1742. https://doi.org/10.1136/bmj.g1742.

Senn, S. (2013). “Seven myths of randomisation in clinical trials.” *Statistics in Medicine*, 32(9): 1439-1450.

Stanhope, J., and Weinstein, P. (2023). “Critical Appraisal in Ecology: What Tools Are Available, and What Is Being Used in Systematic Reviews?” *Research Synthesis Methods* 14(3): 342–56. https://doi.org/10.1002/jrsm.1609.

Sterne, J.A.C. *et al.* (2016). "ROBINS-I: A Tool for Assessing Risk of Bias in Non-Randomised Studies of Interventions." *BMJ* 12: i4919. https://doi.org/10.1136/bmj.i4919.

Sterne, J.A.C. *et. al.* (2019). "RoB 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials." *BMJ* 28: l4898. https://doi.org/10.1136/bmj.l4898.

Valentine, J.C. and Thompson, S.G. (2013). "Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions." *Res. Syn. Meth.* 4: 26-35. https://doi.org/10.1002/jrsm.1064

Wang, Y. et al. (2022). "Instruments Assessing Risk of Bias of Randomized Trials Frequently Included Items That Are Not Addressing Risk of Bias Issues." *Journal of Clinical Epidemiology* 152: 218–25. https://doi.org/10.1016/j.jclinepi.2022.10.018.

Westgate, M.J. (2019). "revtools: An R package to support article screening for evidence synthesis." *Research Synthesis Methods*. doi:10.1002/jrsm.1374.

Westgate, M.J. (2019). "revtools: Tools to Support Evidence Synthesis." R package version 0.4.1, https://CRAN.R-project.org/package=revtools.

Westreich, D., Greenland S. (2013). "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients". *American Journal of Epidemiology* 177(4): 292–298, https://doi.org/10.1093/aje/kws412

Wickham, H. (2023): "stringr", 10.32614/CRAN.package.stringr

Supplementary Material to Culina et. al: Elevating the importance of Risk of Bias assessment for ecology and evolution

1) **Survey on the familiarity with the Risk of Bias**

Welcome to the Survey on Risk of Bias Assessment Tool for evidence synthesis in Ecology & Evolution

This survey is open to researchers in the fields of Ecology and Evolution that have some experience in conducting a systematic review and/or meta-analysis. Its purpose is to inform the development of a risk of bias tool that will be simple and relatively quick to use when assessing bias of studies included in systematic review and meta-analysis in ecology and evolution.

The survey has 12 questions and it takes around 5 minutes to complete. We will randomly select two respondents for a prize, which consists of a PDF version of the Handbook of Meta analysis in Ecology and Evolution (Princeton University press). If you want to be considered in this draw, please leave your email at the end of the survey.

Survey data is collected anonymously. Participant Information Sheet can be found by clicking on this link. The consent is voluntary and can be withdrawn at any time with future effect. If you withdraw your consent, this does not affect the lawfulness of the processing up to that point. (To opt out, please contact the survey contact person - see Participant Information Sheet). In the event that you opt out, your data will be deleted immediately. You will not be put at a disadvantage by refusing or withdrawing your consent. The survey received ethics approval from the Ethics Committee of Ruđer Bošković Institute (Zagreb, Croatia), ref. ZV/3218/1-2023.

Please indicate below if you give consent to take part (YES) or if you would like to leave this survey now (NO).
YES
NO

Survey questions:

1. Prior to receiving this survey invite, were you familiar with the concept of Risk of Bias (RoB)? [Yes/No]
2. If Yes, what do you consider the scope of RoB to be? [free text answer]
3. Are you aware of any Risk of Bias assessment tools (in your or other fields)? [Yes/No].
   If Yes, please list them [free text answer]

4. Have you ever performed a Risk of Bias assessment? [Yes/No]
5. If you have performed a RoB assessment, what tool/approach did you use (including any RoB assessment tool that you may have developed on your own)? [free text answer]
6. If you have performed a RoB assessment, for what purpose (in what context) have you done so? [free text answer]
7. If you have performed a RoB assessment, how easy/difficult would you rate the experience from 1 [very easy] to 10 [very difficult]?
8. If you have performed a RoB assessment, how likely are you to perform one in the future (if the need arises) from 1 [very likely] to 10 [very unlikely]?
9. If you have NEVER performed a RoB assessment, why?
    1) Was not aware it existed
    2) Was aware it existed, but seemed too much work
    3) Was aware it existed, tried to do it, but turned out too difficult
    4) Was aware it existed, but could not find a suitable assessment tool
    5) Other [free text answer]

10. What is your experience with systematic reviews or meta-analyses (multiple options possible):
    (a) I have led one systematic review or meta-analysis
    (b) I have co-authored one systematic review or meta-analysis
    (c) I have led multiple systematic reviews or meta-analyses
    (d) I have co-authored multiple systematic reviews or meta-analyses
    (e) Systematic review or meta-analysis is my specific area of expertise
    (f) None of the above

11. What is your primary field of research: behavioural ecology, evolutionary ecology, plant ecology, animal ecology, population ecology, microbial ecology, aquatic ecology, soil ecology, molecular biology, methodology, other (please specify)

12. At what stage of your research career are you?
    (a) Undergraduate
    (b) Master student
    (c) PhD researcher
    (d) Early-career researcher (i.e., have completed my doctorate degree within the past 5 years)
    (e) Mid-career researcher (i.e. have completed dy doctorate degree 5-10 years ago)
    (f) Senior researcher (i.e., have completed my doctorate degree more than 10 years ago)
    (g) I am not a researcher

If you are interested in receiving a follow-up on this work, or would like to be considered in the prize draw, please email the survey administrator at aculina@irb.hr, indicating the survey

participation and your interests (follow-up, prize draw). Your email will not be used for any other purpose, and will be forgotten after the research is finished. If you wish your email to be forgotten at any time, please email the survey administrator.

—----------

Completion page of survey:

Thank you for completing this survey. Previous research has identified a critical lack of awareness of and engagement with Risk of Bias assessment in Ecology and Evolution (O'Dea et al. 2021), which could be impeding scientific progress in these fields (Gurevtich et al. 2018). We strongly recommend that researchers considering performing a systematic review or meta-analysis familiarise themselves with the concept of Risk of Bias in research synthesis, and explore methods to formally evaluate bias to follow best practice. For an introduction to the topic we suggest:

- O'Dea et al. 2021 Section: 'VIII. PRIMER D: BIAS FROM PRIMARY STUDIES'
- Cooke et al. 2017 Section: 'Evidence synthesis in systematic reviews'.

2) **Email text/ call to participate in the survey**

—----------

Dear colleague,

We would appreciate your help in developing a simple "Risk of Bias" (RoB) assessment tool for ecology and evolution. In many other disciplines, such tools are routinely used for evidence synthesis, including meta-analysis, to qualitatively assess whether biases within reviewed primary studies risk distorting conclusions of the evidence synthesis.

This survey is open to **researchers in the fields of Ecology and Evolution that have some experience in conducting a systematic review or meta-analysis.** We would be grateful if you could spare 5 minutes to fill in an anonymous survey, which can be found here. After filling out the survey, you can opt to be included in the prize draw (PDF version of the Handbook of Meta analysis in Ecology and Evolution , Princeton University Press).

The survey will remain open until the 15th of October 2023. By responding electronically, you will be consenting to have your anonymous responses included in the research project on Risk of Bias tools in EE evidence synthesis.

If you have any further questions about this research or want to have your data removed from the survey, please contact the survey administrator at aculina@irb.hr.

Thank you in advance for your response!

Very Best,
Antica Culina, the survey administrator

3) **Search string** for identifying meta-analyses published accross 300 journals.

(TI=("meta-analy*" OR "metaanaly*" OR "meta-regression" OR "metaregression") OR AB=("meta-analy*" OR "metaanaly*" OR "meta-regression" OR "metaregression") OR AK=("meta-analy*" OR "metaanaly*" OR "meta-regression" OR "metaregression")) AND IS=("13652486" OR "14610248" OR "01678809" OR "16000706" OR "14668238" OR "00063207" OR "14321939" OR "14712954" OR "15231739" OR "20457758" OR "13652664" OR "09218009" OR "13652745" OR "16161599" OR "1365294X" OR "14209101" OR "19395582" OR "13652435" OR "17264189" OR "2041210X" OR "15585646" OR "13652656" OR "21508925" OR "15375323" OR "17517370" OR "16000587" OR "13652699" OR "1744957X" OR "14657279" OR "15731464" OR "15577015" OR "22120416" OR "14350629" OR "15738477" OR "13652427" OR "14712148" OR "15729710" OR "09258574" OR "03043800" OR "14724642" OR "17524571" OR "16541103" OR "1526100X" OR "10958312" OR "15749541" OR "23519894" OR "15735052" OR "01695347" OR "14391791" OR "14320762" OR "21619549" OR "16171381" OR "0022541X" OR "15729761" OR "14338319" OR "14429993" OR "09123814" OR "17083087" OR "17545048" OR "01401963" OR "01692046" OR "15371719" OR "2397334X" OR "14726785" OR "15733017" OR "15409309" OR "1432184X" OR "10968644" OR "1654109X" OR "16161564" OR "19360592" OR "13652540" OR "17550998" OR "00314056" OR "1438390X" OR "15507424" OR "00652504" OR "17596653" OR "10959513" OR "14691795" OR "15891623" OR "20511434" OR "11769343" OR "19342845" OR "22132244" OR "22244662" OR "00220981" OR "14657333" OR "19413300" OR "14697831" OR "01106465" OR "13653008" OR "24501395" OR "15731642" OR "10353712" OR "0003455X" OR "15452069" OR "17447429" OR "11645563" OR "03672530" OR "00472484" OR "17529921" OR "08000395" OR "16083334" OR "1076836X" OR "19385455" OR "13652028" OR "17279380" OR "20412851" OR "10301887" OR "02757540" OR "14230445" OR "10960031" OR "15882756" OR "22145753" OR "15735133" OR "14390574" OR "15220613" OR "19339747" OR "13504509" OR "15731561" OR "14321432" OR "08858608" OR "21996881" OR "14753057" OR "10369872" OR "23524855" OR "02497395" OR "15482324" OR "19968175" OR "18741746" OR "10960325" OR "13991183" OR "19436246" OR "18094392" OR

"1146609X" OR "24108200" OR "00030031" OR "13488570" OR "15735125" OR "18185487" OR "03051978" OR "00030090" OR "0079032X" OR "1065657X" OR "19954263" OR "1432041X" OR "2073106X" OR "16423593" OR "1476945X" OR "14427001" OR "19399170" OR "11956860" OR "20419139" OR "1525142X" OR "21553874" OR "1876312X" OR "03781844" OR "14455226" OR "17513758" OR "00220930" OR "15525015" OR "1944687X" OR "02705060" OR "14645262" OR "14772019" OR "14390469" OR "1860188X" OR "17451019" OR "20513933" OR "00280712" OR "19385307" OR "0029344X" OR "16181077" OR "19385331" OR "0340269X" OR "14421984" OR "14322056" OR "18739652" OR "01380338" OR "19385293" OR "18397263" OR "07176317" OR "19385412" OR "00384909" OR "13653113" OR "05643295" OR "02408759" OR "15270904" OR "09096396" OR "14636409" OR "13541013" OR "1461023X" OR "00301299" OR "1466822X" OR "477525" OR "00298549" OR "09628452" OR "08888892" OR "00218901" OR "00220477" OR "01718630" OR "09621083" OR "1010061X" OR "10510761" OR "02698463" OR "17264170" OR "00143820" OR "00218790" OR "00030147" OR "17517362" OR "09067590" OR "03050270" OR "17449561" OR "10452249" OR "13873547" OR "2818182" OR "14329840" OR "02697653" OR "00465070" OR "09603115" OR "13669516" OR "11009233" OR "10612971" OR "00244066" OR "13850237" OR "03405443" OR "09212973" OR "14429985" OR "07374038" OR "09639292" OR "15409295" OR "00953628" OR "2769667" OR "14022001" OR "09483055" OR "19360584" OR "0018067X" OR "1755098X" OR "14383896" OR "10557903" OR "13679430" OR "00713260" OR "15659801" OR "00221503" OR "00224561" OR "02664674" OR "00306053" OR "15052249" OR "10838155" OR "1543592X" OR "623257" OR "10674136" OR "10635157" OR "00840173" OR "01416707" OR "10220119" OR "09377409" OR "07483007" OR "15858553" OR "22145745" OR "03781909" OR "16124642" OR "00980331" OR "00222844" OR "03782697" OR "00322474" OR "03636445" OR "00400262" OR "18741738" OR "00405809" OR "21933081" OR "02775212" OR "00445967" OR "24107220" OR "09187960" OR "13862588" OR "17986540" OR "19954255" OR "0949944X" OR "2833888" OR "1520541X" OR "19344392" OR "1399560X" OR "15525007" OR "00222933" OR "09475745" OR "18601871" OR "17451000" OR "10926194" OR "14396092" OR "00948373" OR "0913557X" OR "07224060" OR "00973157" OR "0370047X" OR "0716078X" OR "15287092" OR "03076970" OR "03003256")

4) Journals and the Risk of Bias - data extraction form

The authors AC, AST, OP, and ROD have used this form to extract information from journals that publish ecological and evolutionary research, and that related to their reference to Risk of Bias or similar concepts for the authors of evidence synthesis.

Section one

1. 2.Extractors initials [AC, AST, OP, RO]
2. Journal name [free text]
3. Does the journal explicitly solicit some form of evidence synthesis. If you answer NO, submit the form. Otherwise continue to the next section. By evidence synthesis we mean: ' Evidence syntheses are conducted in an unbiased, reproducible way to provide evidence for practice and policy-making, as well as to identify gaps in the research. Evidence syntheses may also include a metaanalysis, a more quantitative process of synthesizing and visualizing data retrieved from various studies' [Yes, No, Unusre]
4. Comments on this section [free text]

Section two: Fill only if the journal explicitly solicits some form of evidence synthesis, or if you are unsure about it.

1. Does journal specifically mention RoB or related assessment of primary literature in guidelines to authors of evidence synthesis, or RoB/related assessment (e.g. critical appraisal, quality assessment) is specifically mentioned in linked other guidelines (e.g. journal states something like ‘follow PRISMA guidelines when reporting MA’ but no further detail on RoB or related is mentioned)
   [Yes, No, Unusre].
2. Comments on this section [free text]

Section three: Fill only if the journal specifically mentions RoB or related assessment in guidelines to authors of evidence synthesis, or RoB/related assessment is specifically mentioned in linked other guidelines.

1. What concept related to RoB (including RoB itself) does the journal or a linked guideline exactly mention? If 'Other', please copy paste the concept. [RoB, Critical appraisal, Qaulity appraisal, Other]

2. If the journal links or refers to external guidelines that mention RoB or related assessment, what are these guidelines. If 'Other', please copy paste their name. [PRISMA, PRISMA-P, PRISMA EcoEvo, CEE guidelines, ROSES, EQUATOR, JBI Guidelines, Does not refer to any guidelines, Other].
3. What specific RoB or related tool/checklist is mentioned in journal guidelines, or in linked guidelines? If Other, please copy paste their name. [No tool mentioned, CEE critical appraisal tool (CEECAT), ROBITT, SYRCLE, Cochrane RoB tool, Cochrane RoB2 tool, Cochrane EPOC checklist, Cochrane ROBINS-I tool, JBI checklist for quasi-experimental studies, Other]
4. Where on the journal webpage is the RoB or related assessment mentioned, or name of external guideline stated, or link to an external guideline provided– copy paste the address. If more separate with ; [free text]
5. What is the strength of the policy on the use of RoB or related? [mandated (wording such as ‘expect’, ‘must’, ‘have to’, require’), encouraged (wording such as ‘should’, recommends’, ‘requests’), optional (wording such as “we invite authors to provide a RoB assessment”, or if they simply mention the guideline but say nothing else")]
6. Comments on this section [free text]