Received: 11 December 2025 | Accepted: 10 June 2026

DOI: 10.1111/2041-210x.70356

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution

REVIEW

Elevating the importance of Risk of Bias assessment for ecology and evolutionary biology

Antica Culina[1] | Dugald Foster[2] | Matthew Grainger[3] | Rose E. O'Dea[4] | Oliver L. Pescott[5] | Alfredo Sánchez-Tójar[6,7,8] | Robin J. Boyd[2,5] | Julia Koricheva[9] | Shinichi Nakagawa[10] | Gavin Stewart[11]

[1]Ruder Boskovic Institute, Zagreb, Croatia; [2]Centre for Ecology and Conservation, University of Exeter, Cornwall, UK; [3]Knowledge Synthesis Department, Norwegian Institute for Nature Research (NINA), Trondheim, Norway; [4]School of Agriculture, Food and Ecosystem Sciences, University of Melbourne, Melbourne, Australia; [5]UK Centre for Ecology & Hydrology, Wallingford, UK; [6]Department of Evolutionary Biology, Bielefeld University, Bielefeld, Germany; [7]CNC-UC, Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal; [8]CIBB, Center for Innovative Biomedicine and Biotechnology, University of Coimbra, Coimbra, Portugal; [9]Department of Biological Sciences, Royal Holloway University of London, Surrey, UK; [10]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada and [11]University of Newcastle, Newcastle upon Tyne, UK

Correspondence
Antica Culina
Email: aculina@irb.hr

Funding information
Fundação para a Ciência e a Tecnologia, Grant/Award Number: LA/P/0058/2020, UID/PRR/4539/2025 and UID/04539; Hrvatska Zaklada za Znanost, Grant/Award Number: HRZZ-IP-2022-10-2872

Handling Editor: Pen-Yuan Hsing

Abstract

1. Systematic reviews and meta-analyses are key evidence synthesis methods for informing future research, interventions and policy. As the validity of their conclusions depends on the primary studies they synthesise, assessing the internal validity of the included studies is essential. In some fields, such as medicine, this is the norm and is commonly done using Risk of Bias assessment tools.
2. Risk of Bias (RoB) assessment is however rare in ecology and evolutionary biology (EEB) even though several RoB tools have been developed in some ecological subfields and related fields. To identify potential reasons for a limited uptake, we conducted a survey of ecologists and evolutionary biologists with evidence synthesis experience and reviewed 275 journals that publish EEB research for guidelines on performing RoB or related assessments.
3. Only 28 of 232 (12%) survey respondents had correct interpretation of the RoB concept, while 46 (40%) of 116 that had heard of RoB have confused RoB with publication bias. Just 10 (4%) had conducted a RoB assessment, most of whom found it challenging. Out of the 209 EEB journals that explicitly solicit evidence synthesis ($N=58$) or reviews ($N=151$), only five (2%) directly mentioned standards for conducting evidence synthesis, which include RoB or related (e.g. critical appraisal) assessments. An additional 45 (22%) journals indirectly linked to RoB or a related assessment via referring to the guidelines for reporting evidence synthesis (e.g. PRISMA), despite such reporting guidelines not providing information on how to conduct RoB assessments.

4. To increase its uptake in EEB we recommend making RoB assessment: (1) known and recognised as an essential component of a reliable evidence synthesis by including it in training materials, and journals' and funders' guidelines and policies; (2) easy to perform by bringing the synthesis community together to determine the need for developing new or adjusting existing RoB tools; and (3) possible by further improving reporting standards for primary studies so that RoB assessment can be done on these studies. For those unfamiliar with the RoB assessment, we provide five key RoB questions that existing tools often cover. These questions can be considered to understand the basic composition of the evidence included in evidence synthesis.

KEYWORDS
best practices, bias assessment, evidence-base, meta-research, reproducibility, research integrity, research synthesis, study quality

BOX 1 Aspects that affect trust in primary studies and in the overall evidence synthesis.

Primary study (adapted from Juni, 2001; Guyatt et al., 2011; Sedgwick, 2014; Frampton et al., 2022; 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). Internal validity will depend on the extent of any trend or systematic deviation in data collection, data analysis or data interpretation. The potential threats to the internal validity are called the Risk of Bias.

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).

External validity—the extent to which results of a study can provide a correct basis for generalisation to other circumstances (i.e. generalised to the population that the sample was meant to represent).

Evidence synthesis:

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 unconsciously exclude studies that do not support such an opinion. See, for example, Burghardt et al. (2012), Bishop (2017) and Ciria et al. (2023).

Literature bias—a set of biases present in the overall set 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. (2020) and Nuñez and 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 literature bias). See Haddaway and Verhoeven (2015) and 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 and Macura (2018) and O'Dea et al. (2021).

1 | 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 evolutionary biology can be investigated by conducting experimental or observational studies (i.e. primary studies). However, the findings of primary studies are mostly limited to an often narrow context, and generalisation frequently requires synthesis. 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. In evidence synthesis, information on a topic of interest is systematically searched for and reviewed, commonly via systematic review and often subsequently quantitatively summarised using meta-analysis (Arnqvist & Wooster, 1995).

Evidence synthesis is becoming essential to inform research and policy development and to 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 (Grainger et al., 2020; Nakagawa et al., 2020). The robustness of the evidence synthesis, and thus its conclusions, depends on how evidence synthesis is conducted and reported (summarised in Box 1). Thus, it is of the utmost importance that systematic reviews and meta-analyses are conducted and reported according to the highest standards. Systematic reviews and meta-analyses in ecology and evolutionary biology have adopted more rigorous methodological approaches and reporting standards over time (Koricheva & Gurevitch, 2014; Lodi et al., 2021). Many of these were taken and adjusted from medicine, a field where reliable systematic reviews and meta-analyses have direct implications for human health interventions. For example, the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines and checklist (Moher et al., 2009; Page et al., 2021) were recently extended to PRISMA-EcoEvo (O'Dea et al., 2021), and the ROSES standards (RepOrting standards for Systematic Evidence Syntheses in environmental research, Haddaway et al., 2018) were recently developed. Methodological approaches to testing for publication bias in ecological and evolutionary meta-analyses have also been recently reviewed and extended (Nakagawa et al., 2022).

Publication bias and other sensitivity analyses, and reporting following reporting standards take place after evidence has already been included in a meta-analysis. What is typically missing is a step that clearly and transparently evaluates the evidence before summarising it, that is, an assessment of the internal validity of the primary studies included (see Box 1). Internal validity describes the extent of systematic (i.e. non-random) error inherent to an individual study (Frampton et al., 2022). If we cannot trust some primary studies, can we trust the evidence synthesis including these studies?

The potential threats to the internal validity of a study are often termed 'Risk of Bias' (RoB), and are commonly assessed using RoB tools. Any 'systematic deviation from the truth', that is, any systematic bias (e.g. Wang et al., 2022) might lead to invalid conclusions of a primary study, but also of a synthesis of such studies. We note that the general phrase 'Risk of Bias' is sometimes used to describe other aspects related to bias, usually a bias within the overall pool of studies relative to some larger inferential population (see Box 1), such as geographic (e.g. Martin et al., 2012), language (e.g. Nuñez & Amano, 2021), taxonomic (e.g. Troudet et al., 2017) or publication bias (e.g. Nakagawa et al., 2022), which can also lead to further confusion, as we discuss later. Here, we focus on within-study RoB and argue that assessing the RoB of each included primary study (i.e. assessing its internal validity) should be standard practice in ecological and evolutionary evidence synthesis. Moreover, beyond the context of evidence synthesis, researchers can use RoB tools to help minimise bias when designing and conducting primary studies.

While RoB tools have been developed in the field of environmental evidence, which intersects with ecology (Konno et al., 2021), for studies assessing temporal trends in ecology (Boyd et al., 2022) and for animal studies (Hooijmans et al., 2014; Sterne et al., 2016), these tools have not become widely used in ecology and evolutionary biology (Stanhope & Weinstein, 2023). Further, while some reporting standards that have increasingly been adopted in meta-analysis (such as PRISMA-EcoEvo) mention RoB or related concepts, they do not provide guidance on how to conduct RoB assessments.

With this contribution, we aim to raise awareness of and promote the use of RoB assessment in ecology and evolutionary biology by: (1) stressing the importance of performing RoB assessments in evidence synthesis; (2) understanding the reasons for the low prevalence of RoB assessments in these fields by surveying evidence synthesists and journals; (3) suggesting key actions to promote RoB assessments. We also summarise the main elements of several existing RoB tools into five questions that readers could consider when conducting evidence synthesis or designing primary studies. We argue that applying the RoB assessment approach to both the design and evaluation of primary studies can reduce bias and strengthen the inferences we can draw from ecological and evolutionary primary studies.

1.1 | The importance of Risk of Bias assessment

In this section, we highlight the importance of RoB assessments for readers who are not familiar with the concept. We adopted the common view that RoB pertains to the internal validity of the study (Wang et al., 2022; Frampton et al., 2022; but see Pescott et al., 2023). A study's internal validity describes the extent of its systematic error—that is, the extent to which the expected value of the within-study estimate differs from the within-study target parameter owing to flaws in study design, conduct or analysis rather than random error (Frampton et al., 2022). This may be contrasted with external validity, which concerns the extent to which the within-study target parameter can be used to answer a broader question without systematic error (Frampton et al., 2022). Internal validity may be compromised by confounding, which occurs when a common cause of the explanatory variable and the response induces a non-causal dependence between the two. Observational studies, which do not randomise treatment assignment, are particularly susceptible to confounding. Another threat to internal validity is 'structural' selection bias, which occurs when selection of units into the sample creates a non-causal dependence between the explanatory variable and the response (Hernán et al., 2004).

The RoB assessment can be a component of a broader assessment of other aspects at the level of the primary study (e.g. external validity, reporting quality, precision; see Box 1), or biases at the level of the overall evidence synthesis (e.g. reviewer/researcher bias, taxonomic bias; see Box 1). Unfortunately, the literature is frequently unclear or inconsistent on the definitions of the above-mentioned terms, which likely contributes to the confusion about their application and intent (cf. Wang et al., 2022). For example, the term 'critical appraisal' can sometimes mean assessing RoB only, but in other cases can include other aspects, such as transparency of primary studies (see examples in Stanhope & Weinstein, 2023).

The internal validity of a study will depend on the presence of biases in data collection, analysis, and results interpretation. These 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 considering those potential biases. Neglecting RoB in evidence synthesis can lead to incorrect conclusions. Meta-analyses including studies at higher Risk of Bias can systematically overestimate treatment effects, and restricting to low-risk studies can substantially reduce effect sizes and, in many cases, alter or reverse conclusions (e.g. Pildal et al., 2007). Another example is early pooled summaries and published trial reports that supported the idea that antiviral medication oseltamivir reduced serious complications and justified large government stockpiles (e.g. Kaiser et al., 2003). An updated Cochrane systematic review obtained full clinical study reports, applied rigorous RoB assessment and re-analysed the trials and found both far weaker evidence of benefit and nontrivial harms (Jefferson et al., 2014), casting doubt on the rationale for the huge stockpiling policies implemented between 2009 and 2010. In a field that intersects with ecology, Konno et al. (2025) collected articles that assessed the impact of different biases (including what we consider here under the Risk of Bias) on effect estimates in environmental research, finding that some types of bias influenced effect size estimates (e.g. confounding bias, detection bias, selection bias).

RoB tools have been developed to facilitate and standardise RoB assessment and are generally required and extensively used in medicine (especially for clinical trials). For example, out of 1172 randomised clinical trials systematic reviews that were published between 2023 and 2024, 1140 (97.3%) reported assessing the quality or RoB of the primary studies (Sandoval-Lentisco et al., 2024). Relatively close to ecology, the Center for Environmental Evidence has been developing a CEECAT tool (Collaboration for Environmental Evidence Critical Appraisal Tool; Konno et al., 2021). Similar tools exist for studies assessing temporal trends in ecology (Boyd et al., 2022) and for animal studies (Hooijmans et al., 2014; Sterne et al., 2016). However, these seem not to be widely used as a recent analysis of 680 systematic reviews in ecology found that only 4% included any form of critical appraisal (Stanhope & Weinstein, 2023).

1.2 | Understanding the reasons behind the low use of RoB assessments

Given their importance for the validity of the conclusions 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 reported having published systematic reviews and meta-analyses. In addition, we reviewed the guidelines of 275 ecology and evolution journals.

2 | METHODS

2.1 | Survey on 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 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 RoB, awareness and use of RoB assessments and questions on familiarity with meta-analysis, their field of research and career stage (survey questions can be found in the Supporting Information).

The survey, created in Google Forms, was sent on the 11th September 2023 to the emails of 3346 corresponding authors of meta-analyses in ecology and evolutionary biology, as well as to several mailing lists (NC^3 Collaborative Research Centre; Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology; German Zoological Society; Sociedad Española de Etología y Ecología Evolutiva; europelist@conbio.org). The survey was also advertised via Slack channels (Big Team Science Conference Slack Channel, German Reproducibility Network Event Slack Channel, ESMARConf Slack Channel), Twitter, and the SORTEE newsletter. The survey was open to responses until the 15th October 2023. The invitation email text is available in the Supporting Information.

To determine the corresponding authors of published meta-analyses, we searched for meta-analyses published in 300 journals that publish ecology and evolutionary biology research via Web of Science (databases: SCI-EXPANDED, SSCI, AHCI, ESCI) on the 25th April 2023 (the search string, with journal codes, is available in the

Supporting Information and in the Culina et al., 2025) (datasets and codes for this work). The search retrieved 3289 results (potential meta-analyses), BibTeX list of these available in the datafile (Culina et al., 2025) published between 1991 and 2023. We then extracted a list of all corresponding email addresses from these articles using packages revtools 0.4.1 (Westgate, 2019a, 2019b) and stringr 1.5.0 (Wickham, 2023) in R 4.2.3 (R Core Team, 2021). Code is available in the data file (Culina et al., 2025). This resulted in 3346 email addresses we sent the email to, of which 789 (~24%) bounced back.

The survey received 232 responses. Most of the respondents were 10+ years after their PhD ($n=106$). Others were 5–10 years ($n=57$) and within 5-year post-PhD ($n=40$), PhD student ($n=22$), undergraduate ($N=3$), non-researcher ($N=3$) and master student ($N=1$). Respondents were relatively experienced in conducting systematic reviews and meta-analysis: 73 had led several such studies, 63 had co-authored multiple, and 87 had led one, while 24 had co-authored one meta-analysis (note that there is a minor overlap between these groups). Details on career stage, expertise level and research area of the respondents can be found in the data file (Culina et al., 2025).

Some responses were prone to subjective judgement regarding their meaning. Thus, four assessors (AC, AST, MG, REO; 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 for each of those 188 respondents:

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. Does the respondent think RoB is publication bias?
6. Has the respondent likely conducted publication bias, rather than RoB assessment?

We then compared the responses among all four assessors. When more than two assessors had the same interpretation, we chose the most common answer as the final one. For disagreements (26 out of 708 questions), 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 assessor and the table with the final (combined) scores are available in data file (Culina et al., 2025).

2.2 | Journals and RoB assessment

Between 19th April 2025 and 12th May 2025, authors AC, AST, OLP and REO checked the websites of 275 ecology and evolutionary biology journals. The list of journals was taken from Ivimey-Cook et al. (2025) and is available from the data file (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 RoB or related concepts (e.g. critical appraisal) for authors of evidence synthesis articles. Here, a journal can explicitly mention a need for a RoB or a related assessment (e.g. critical appraisal), it can link to the external guidelines on RoB or a related assessment (e.g. CEE guidelines, JBI guidelines on critical appraisal), or it can link to reporting guidelines that just mention that RoB (or related) assessment should be reported (e.g. PRISMA, PRISMA-EcoEvo, ROSES).

We piloted the data extraction on 10 journals to adjust the data extraction questions and align responses between the reviewers. We used Google Forms for data extraction. The exact extraction questions can be found in the Supporting Information. The main questions, with the potential answers in brackets, were as follows:

1. Does the journal explicitly solicit some form of evidence synthesis? [Yes, No, Unsure]
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)? [Yes, No, Unsure]
3. What concept related to RoB (including RoB itself) does the journal or a linked guideline exactly mention? [RoB, Critical appraisal, Quality appraisal, Other]
4. If the journal links or refers to external guidelines that mention RoB or related assessment, what are these guidelines? [PRISMA, PRISMA-P, PRISMA-EcoEvo, CEE guidelines, ROSES, EQUATOR, JBI Guidelines, Does not refer to any guidelines, Other].
5. What specific RoB or related tool/checklist is mentioned in journal guidelines, or in linked guidelines? [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]
6. What is the strength of the journal's policy on the use of RoB or related assessment? [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')]

Based on the pilot extraction, we decided to: (a) follow 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); (b) score 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) 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 data file (Culina et al., 2025).

3 | RESULTS

3.1 | Survey on RoB awareness and use

Only 28 of 232 (12%) survey respondents correctly interpreted RoB (Figure 1), and only 10 (4%) had used a RoB tool (an existing one, or one they had developed themselves). While around half of our survey respondents ($N=116$) had heard of RoB, it appears that many confused RoB with publication bias ($N=46$, Figure 1) or had some other interpretation. Similarly, while 66 respondents claimed to be aware of some RoB tools, when asked to list some, 37 listed publication bias tools. Finally, 53 respondents answered that they had conducted a RoB assessment; however, 34 had actually used some form of publication bias test (Figure 2).

Out of 10 respondents who had conducted RoB assessment, 8 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.

3.2 | Journals and RoB assessment

Out of 275 journals, only 58 explicitly solicited evidence synthesis, while 151 solicited the vague term 'Reviews'. Out of these 209 journals that solicited evidence synthesis or reviews, only five explicitly mentioned standards of conducting evidence synthesis, which include RoB or a related (e.g. critical appraisal) assessment (Figure 3). These journals either then linked 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 did not mention standards for conducting evidence synthesis, but rather reporting guidelines for evidence synthesis (e.g. PRISMA) and through these reporting guidelines only indirectly (and likely unintentionally) link to RoB assessment.

4 | DISCUSSION

4.1 | Increasing the application of RoB assessment in ecology and evolutionary biology

RoB tools developed in other fields (including those intersecting with ecology) are rarely used in ecology and evolutionary biology (Stanhope & Weinstein, 2023). Based on the findings of our 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 (Figure 4).

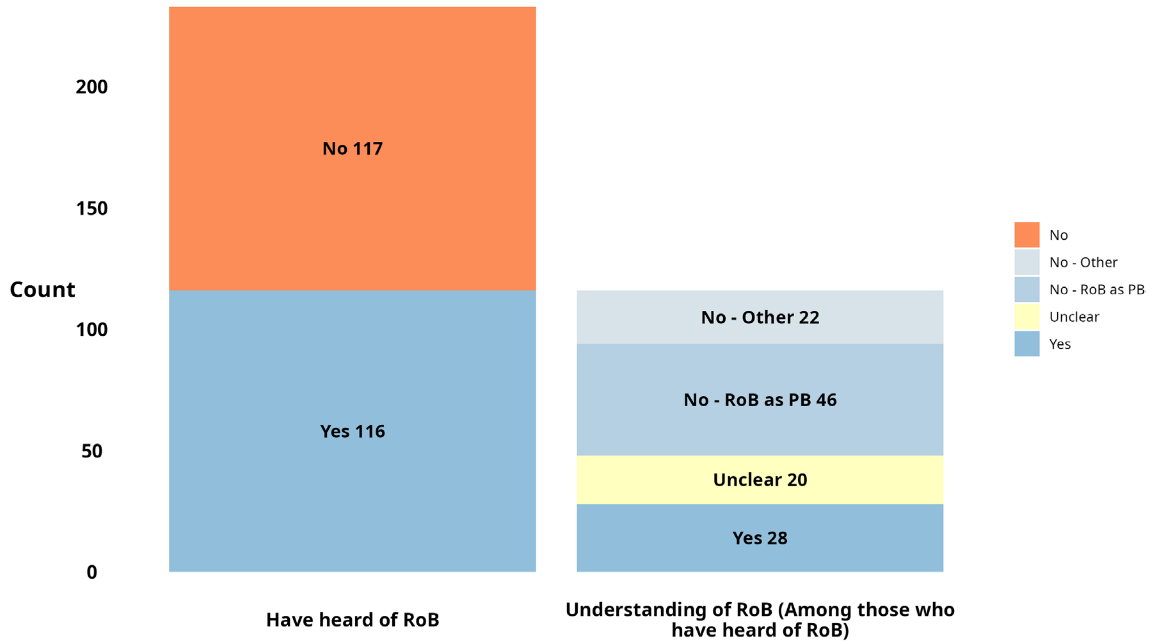

FIGURE 1 Survey results on RoB awareness among ecologists and evolutionary biologists with experience in evidence synthesis. Left-hand bar: Respondents who claimed to have heard of RoB ('Yes') or not ('No'). Right-hand bar: Respondents who had a correct understanding of RoB ('Yes'), those who confused 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 the data file (Culina et al., 2025).

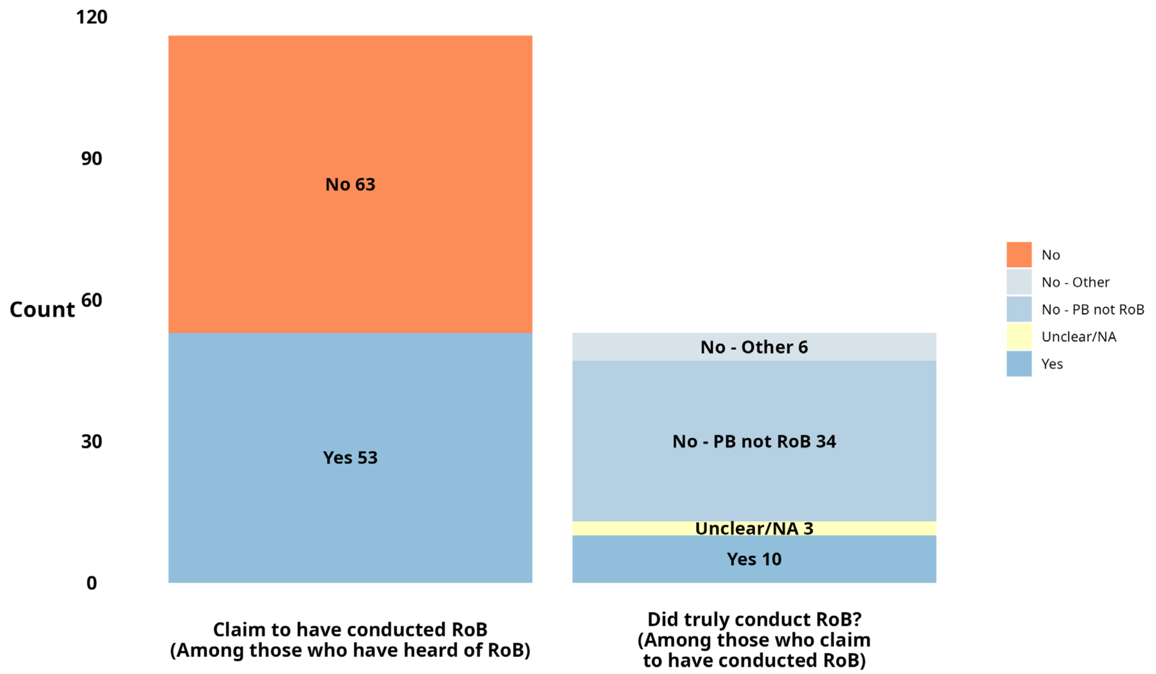

FIGURE 2 Survey results on experience in conducting RoB assessment among ecologists and evolutionary biologists who were aware of RoB. Left-hand bar: The respondents who claimed to have conducted a RoB assessment ('Yes') or not ('No'). Right-hand bar: The respondents who truly had conducted a RoB assessment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

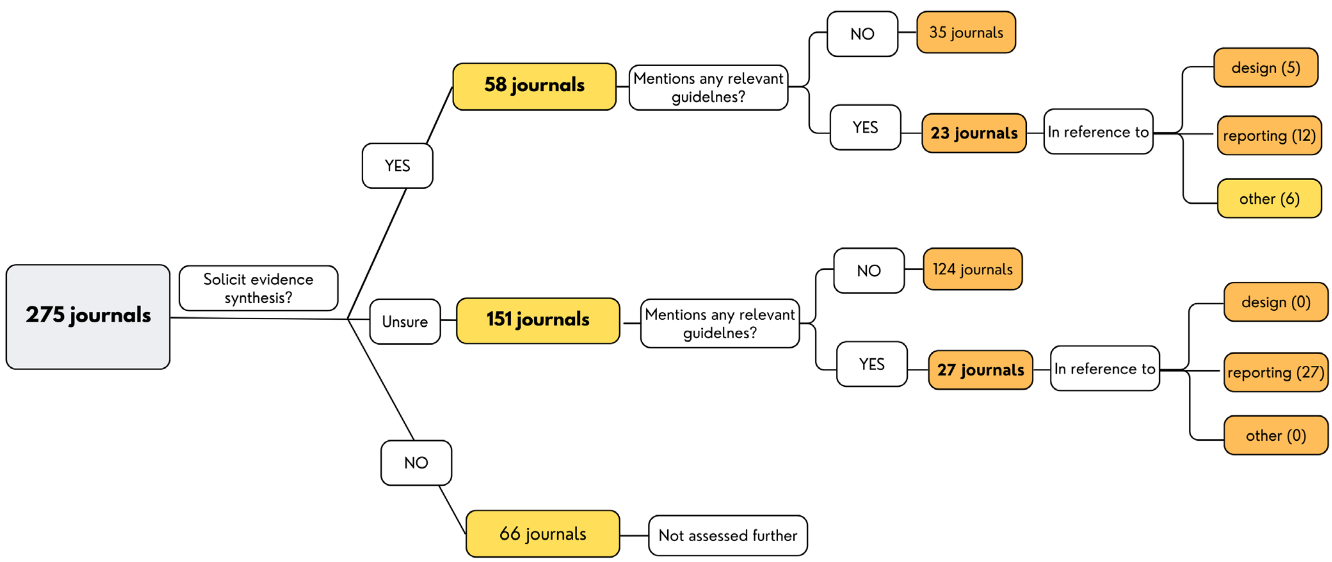

FIGURE 3 Summary results on the mention of RoB or related assessment by 275 journals that publish ecological and evolutionary biology research. We first assessed whether journals solicited evidence synthesis ('YES') or just vaguely referred to 'Reviews' ('Unsure'). We then checked if they mentioned any guidelines 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. Bold text represents journals that were assessed in the next steps.

4.1.1 | Raising awareness and providing guidelines

Our survey results indicated that the primary driver of the low uptake of RoB assessment might be low awareness of its existence and importance among evidence synthesis researchers and journals. To improve this situation, we recommend that RoB assessments are as follows: included in evidence synthesis teaching materials; explicitly mentioned in journals' and funders' guidelines and requirements for evidence synthesis; and promoted via editorials, perspectives and similar articles. We also noticed that most of the journals (151 of

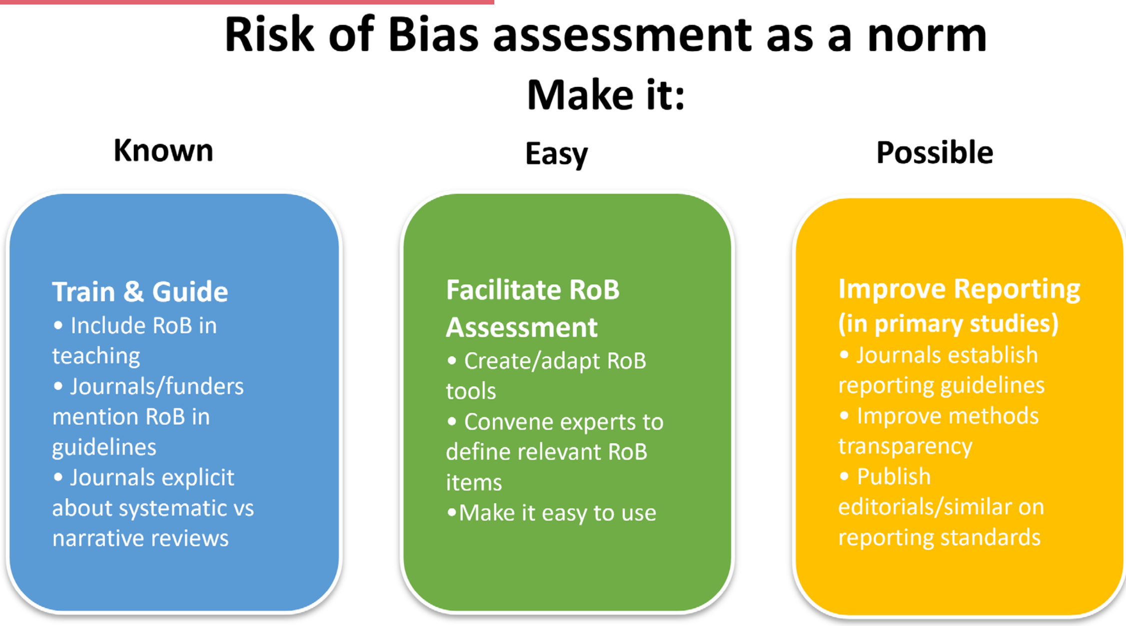

FIGURE 4 The main sets of actions, with examples, to increase the use of Risk of Bias (RoB) assessment in ecology and evolutionary biology.

275) just mention accepting 'Reviews', without further specifying whether these include systematic reviews or only narrative reviews. Being explicit about the types of evidence synthesis, together with clear guidelines on conducting and reporting them, would be an important step in ensuring high-quality evidence synthesis in ecology and evolutionary biology.

4.1.2 | Facilitating RoB assessment

Another reason behind the low update of RoB assessment could be that RoB tools are perceived as too time-consuming to apply. Our survey indicates that the 10 authors who had conducted RoB assessment found it relatively difficult. 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 min for assessing each study. Since the average meta-analysis in ecology and evolutionary biology includes about 40 studies (median = 41 studies in plant ecology, Koricheva & Gurevitch, 2014; average = 40 studies in EEB, Yang et al., 2025), that could constitute an average of around 240 h of additional work per meta-analysis, or 30 8-h days. While large centres for evidence synthesis (e.g. Cochrane, Campbell) have large teams dedicated to and trained for conducting evidence synthesis (including RoB assessment), most ecological and evolutionary synthesis work is done by less specialised and smaller teams. Small team size combined with often non-tailored yet detailed tools might further lead to low use of RoB assessments in ecology and evolutionary biology.

Adjusting the existing tools or potentially creating new ones that would fit the requirements of a wide range of research studies in ecology and evolutionary biology is challenging. In Box 2, we provide the main steps in creating a robust RoB tool, and hope that this article will encourage ecologists and evolutionary biologists to embark on this critical step or to at least reach a consensus that one of the existing tools should be used and potentially adjusted.

4.1.3 | Making RoB possible—reporting standards

While we did not ask this question in the survey, the well-known low reporting quality of primary studies will make 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 et al., 2016). Such poor reporting makes RoB assessment difficult or impossible and can also lead to exclusion of primary studies from meta-analyses (e.g. Moran et al., 2021; Santostefano et al., 2025). Often, 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 (e.g. such as in CEECAT; Konno et al., 2021).

Journals play a crucial role in establishing detailed reporting guidelines of published articles. For instance, to improve the transparency of reporting and the reproducibility of published results, authors of research papers in life sciences, behavioural and social sciences, and ecology, evolution and 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*, which in 2012

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. Boyd et al., 2022; Higgins et al., 2011). 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, 2009; Hartling et al., 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 & Higgins, 2016) and a separate systematic review of the 'theoretical and conceptual literature on types of bias in epidemiology' (Sterne 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**)?

introduced reporting standards in experimental studies (Hillebrand & Gurevitch, 2013), and *Functional Ecology* that has recently introduced the requirement of a 'replication statement' to be included in the Materials and Methods section. A growing number of contributions explain how to report the primary studies so they can be easily included in evidence synthesis (e.g. Gerstner et al., 2017; Hennessy et al., 2022). However, requiring reporting guidelines or providing reporting checklists must be accompanied by other actions that ensure that these are followed.

4.2 | Assessing RoB—A basic guideline

Many existing, detailed and validated tools could be consulted, adjusted or directly 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 (see in Stanhope & Weinstein, 2023). 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 the internal validity of the study.

For those not familiar with existing RoB tools, we provide a list of five questions that summarise the main items of RoB assessment. We created these based on the existing RoB assessment tools (CEECAT, Konno et al., 2021; ROBINS-I, Sterne et al., 2016) and the Catalogue of Bias (https://catalogofbias.org/). While we recommend the use of some of the existing, tested and well-developed tools, researchers can use our questions to understand the basic composition of the evidence they include in meta-analysis or to evaluate if the design of the primary study they plan might have some issues. These questions are as follows:

4.2.1 | Was randomisation applied? (relevant mostly for experimental studies)

In experiments, randomisation involves assigning experimental units (e.g. individuals, plots) randomly to each treatment. Randomisation minimises any differences between the experimental units in the control and treatment groups (but see Senn, 2013) that do not come from the difference in the treatment. A general overview of the importance of randomisation in inferring causality can be found in Greenland (1990).

In observational studies, units are not assigned to exposure by the investigator; instead, inferring the causality depends on sampling design or statistical strategies to approximate random assignment. 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 this case, analytical approaches might partly account for the issue, although inferential certainty might be weakened (e.g. yielding lower precision).

4.2.2 | Was selection bias avoided?

Selection bias refers to situations where study units have unequal chances to be included in the sample depending on a trait of interest that the study investigates. This type of bias is often present in observational studies and experimental studies that have failed to properly randomise their control and experimental groups.

For example, Culina et al. (2013) were interested in whether individuals in pairs of great tits (*Parus major*) that remained together over the successive breeding seasons had higher over-winter survival than birds from those pairs that have changed partners. Here, assigning a bird as alive or dead in each year depended on whether a bird was recaptured in the wild. For both sexes, Culina et al. (2013) found that females and males that remained with their partner were more likely to be recaptured than were the individuals that had changed partners. Thus, the subsequent survival model has accounted for this difference in recapture rate. Not doing so would falsely lead to lower estimates of survival for the individuals that have changed partners.

Selection bias can also happen when the method of selecting study units leads to a sample that does not represent the intended study population (this may also be referred to as sampling bias). For example, this happens when a study uses only adult animals but makes inferences about the population of any age. In this case selection bias will influence the external validity (see Box 1) of the study, but not its internal validity, and thus RoB assessment.

4.2.3 | Was confounding considered? (relevant for observational studies, and experiments that failed to properly randomise)

Biases can arise due to an uncontrolled (or inappropriately controlled) variable (confounder) that influences both the explanatory variable and the outcome (more in Marshall, 2024); including studies where no or an inappropriate control group is used (e.g. Engqvist & Reinhold, 2016).

Byrnes and Dee (2025) provide an example where researchers wanted to study the causal effect of temperature on snail abundance in the Gulf of Maine. For this, they measured snail abundance and temperature at several locations. However, snail abundance is also driven by snail recruitment that partly depends on the regional oceanography. Thus, oceanography influences both water temperature and recruitment. Here, recruitment and oceanography are unobserved confounding variables. Byrnes and Dee (2025) used simulated data to compare the performance of different types of modelling approaches and showed that simple regression models (the type ecologists would typically use) consistently produce estimates that are well-below the true effect size.

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 primary studies 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, it might not be possible to include such 'diverse' effect sizes in the meta-analysis. Thus, combining results from ecological and evolutionary biology studies that use either inappropriate controls for confounders, or else different sets of controls may impact the overall reliability of the meta-analytic estimates (Valentine & Thompson, 2013). Confounding bias can also occur when meta-analyses include primary study estimates for the effects of variables originally modelled as controls. A nice example can be found in Westreich and Greenland (2013) who demonstrate how this can lead to confusion between direct-effect estimates and total-effect estimates for covariates in the model. Petersen et al. (2022) and Konno et al. (2021) provide tools and further details on assessing confounding bias in systematic reviews of observational studies.

4.2.4 | Were the methods for obtaining or 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, a study wants to examine if male birds have larger bills than females, using the same measurement method. The measurement in females is always taken by observer A, and the measurement in males 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. For example, Clark et al. (1991) found that there was a statistically significant interobserver effect in the American crow (*Corvus brachyrhynchos*) beak width measurements.

An outcome can also be measured in different ways. For example, collecting bird sperm for measuring sperm length can be done using either faecal samples or abdominal massage. Girndt et al. (2017) demonstrated that sperm was significantly shorter in faecal samples compared to abdominal massage samples.

4.2.5 | Was the observer or analyst blind to the allocation of subjects?

Blinding is a procedure where a person collecting or analysing data (or subjects of data collection in human studies) is unaware of any information that could bias observations (e.g. awareness of the treatment condition of a sample). 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 (Holman 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 & Perlmutter, 2015).

As an example, Kozlov and Zvereva (2015) asked 31 researchers to measure asymmetry of 100 leaves (scanned images) collected from the same tree. The researchers, experienced in studying fluctuating asymmetry, were told that the study they were participating in was testing the working hypothesis that the asymmetry was larger in highly polluted areas. The researchers were either provided (i.e. were not blind to the 'treatment') or not provided (i.e. were blind to the 'treatment') with the information on the area that the leaves supposedly came from. The researchers who thought that they were handling leaves from a heavily polluted area reported significantly higher values of FA compared to the researchers who were told that the leaves came from an unpolluted site. This clearly demonstrates that having information on the sample can influence the judgement of the experts.

5 | LIMITATIONS AND FUTURE WORK

Our work has several limitations. First, we assume that assessing and accounting for RoB of bias in primary studies will yield less biased meta-analysis estimates in ecology and evolutionary biology. However, this assumption cannot be tested as an insufficient number of meta-analyses in these fields include a RoB assessment. If the use of RoB assessment would increase, it would become possible to evaluate its impacts on meta-analytic results. In a field close to ecology, Konno et al. (2025) compiled papers that evaluated the potential impacts of 39 types of bias (including RoB as defined in our work) in environmental research, and their results indicate that different biases could change the estimated effects.

Second, our survey respondents might not be representative of a larger population of evidence synthesists in ecology and evolutionary biology. However, our estimates of the awareness and use of RoB were so low that we doubt that the overall population would show anything near adequate rates. Further, our estimate of how many researchers did apply RoB (around 4.3%) is almost identical to the rates of use of any form of critical appraisal in systematic reviews in ecology, which was 4% (Stanhope & Weinstein, 2023).

6 | CONCLUSIONS

Risk of Bias (RoB) assessment is essential for the transparency and quality of evidence synthesis results, and it enables explorations of effect heterogeneity and reliability. Our work proposes several main steps towards increasing the use of RoB assessment in ecology and evolutionary biology and highlights the important role of the synthesis community, but also of journals and funders. It is on the community to start teaching and applying RoB assessment and discussing the need for adjustment of the existing tools, or the development of new ones. It is on journals and funders to provide clear guidance on the highest quality standards by explicitly linking to existing best practice for conducting and reporting evidence synthesis. Journals should also clearly state whether they accept systematic reviews, meta-analyses and other evidence synthesis contributions in their 'Aims & Scope' or under 'Article types'. Journals should further aim to improve the reporting quality of primary studies (as previously recommended, e.g. Ihle et al., 2017; Gerstner et al., 2017; Hennessy et al., 2022), which would enable easier assessment of RoB. 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.

RoB assessment should 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 promote the use of the Risk of Bias assessment in ecological and evolutionary synthesis. We believe our contribution could be a starting point for this positive change.

AUTHOR CONTRIBUTIONS

Conceptualisation, data curation, and funding acquisition: Antica Culina; investigation: Antica Culina, Matthew Grainger, Rose E. O'Dea, Oliver L. Pescott, Alfredo Sánchez-Tójar; methodology: all authors; project administration: Antica Culina; software: Antica Culina, Alfredo Sánchez-Tójar; supervision: Antica Culina; visualisation: Antica Culina; writing—original draft: Antica Culina; writing—review and editing: all authors.

ACKNOWLEDGEMENTS

We are grateful to Dr. Danijel Žanko for assisting in creating Figures 1 and 2.

CONFLICT OF INTEREST STATEMENT

Robin Boyd is an Associate Editor at Methods in Ecology and Evolution, but played no role in the handling or reviewing of this manuscript. We declare no other conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70356.

DATA AVAILABILITY STATEMENT

Data and scripts used in this study are available via https://doi.org/10.5281/zenodo.15630232 as 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.

ORCID

Antica Culina https://orcid.org/0000-0003-2910-8085
Dugald Foster https://orcid.org/0000-0002-0727-830X
Matthew Grainger https://orcid.org/0000-0001-8426-6495
Rose E. O'Dea https://orcid.org/0000-0001-8177-5075
Oliver L. Pescott https://orcid.org/0000-0002-0685-8046
Alfredo Sánchez-Tójar https://orcid.org/0000-0002-2886-0649
Robin J. Boyd https://orcid.org/0000-0002-7973-9865
Julia Koricheva https://orcid.org/0000-0002-9033-0171
Shinichi Nakagawa https://orcid.org/0000-0002-7765-5182
Gavin Stewart https://orcid.org/0000-0001-5684-1544

REFERENCES

Arnqvist, G., & Wooster, D. (1995). Meta-analysis: Synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution*, *10*, 236–240.

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., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin, G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J., & Pescott, O. L. (2022). ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. *Methods in Ecology and Evolution*, *13*, 1497–1507. https://doi.org/10.1111/2041-210X.13857

Burghardt, G. M., Bartmess-LeVasseur, J. N., Browning, S. A., Morrison, K. E., Stec, C. L., Zachau, C. E., & Freeberg, T. M. (2012). Perspectives—Minimizing observer bias in behavioral studies: A review and recommendations. *Ethology*, *118*, 511–517. https://doi.org/10.1111/j.1439-0310.2012.02040.x

Byrnes, J. E. K., & Dee, L. E. (2025). Causal inference with observational data and unobserved confounding variables. *Ecology Letters*, *28*, e70023. https://doi.org/10.1111/ele.70023

Ciria, L. F., Román-Caballero, R., Vadillo, M. A., Holgado, D., Luque-Casado, A., Perakakis, P., & Sanabria, D. (2023). An umbrella review of randomized control trials on the effects of physical exercise on cognition. *Nature Human Behaviour*, *7*, 928–941. https://doi.org/10.1038/s41562-023-01554-4

Clark, R. G., James, P. C., & Morari, J. B. (1991). Sexing adult and yearling American crows by external measurements and discriminant analysis. *Journal of Field Ornithology*, *62*, 132–138.

Cooke, S. J., Birnie-Gauvin, K., Lennox, R. J., Taylor, J. J., Rytwinski, T., Rummer, J. L., Franklin, C. E., Bennett, J. R., & Haddaway, N. R. (2017). How experimental biology and ecology can support evidence-based decision-making in conservation: Avoiding pitfalls and enabling application. *Conservation Physiology*, *5*, cox043. https://doi.org/10.1093/conphys/cox043

Crocker, T. F., Lama, L., Jordãoa, M., Brundlea, C., Prescotta, M., Forstera, A., Ensorb, J., Gladmanc, J., & Clegga, A. (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., & Sheldon, B. C. (2013). A multievent approach to estimating pair Fidelity and heterogeneity in state transitions. *Ecology and Evolution*, *3*, 4326–4338. 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

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

Foo, Y. Z., O'Dea, R. E., Koricheva, J., Nakagawa, S., & 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*, 1705–1720. https://doi.org/10.1111/2041-210X.13654

Frampton, G., Whaley, P., Bennett, M., Bilotta, G., Dorne, J. L. C. M., Eales, J., James, K., Kohl, C., Land, M., Livoreil, B., Makowski, D., Muchiri, E., Petrokofsky, G., Randall, N., & Schofield, K. (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., Moreno-Mateos, D., Gurevitch, J., Beckmann, M., Kambach, S., Jones, H. P., & Seppelt, R. (2017). Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods in Ecology and Evolution*, *8*, 777–784. https://doi.org/10.1111/2041-210X.12758

Girndt, A., Cockburn, G., Sánchez-Tójar, A., Løvlie, H., & Schroeder, J. (2017). Method matters: Experimental evidence for shorter avian sperm in faecal compared to abdominal massage samples. *PLoS One*, *12*(8), e0182853. https://doi.org/10.1371/journal.pone.0182853

Grainger, M. J., Bolam, F. C., Stewart, G. B., & Nilsen, E. B. (2020). Evidence synthesis for tackling research waste. *Nature Ecology & Evolution*, *4*, 495–497. https://doi.org/10.1038/s41559-020-1141-6

Greenland, S. (1990). Randomization, statistics, and causal inference. *Epidemiology*, *1*, 421–429.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Falck-Ytter, Y., Jaeschke, R., Vist, G., Akl, E.

A., Post, P. N., Norris, S., Meerpohl, J., Shukla, V. K., Nasser, M., Schünemann, H. J., & The GRADE Working Group. (2011). GRADE guidelines: 8. Rating the quality of evidence—Indirectness. *Journal of Clinical Epidemiology*, *64*, 1303–1310. https://doi.org/10.1016/j.jclinepi.2011.04.014

Haddaway, N. R., & Macura, B. (2018). The role of reporting standards in producing robust literature reviews. *Nature Climate Change*, *8*, 444–447. https://doi.org/10.1038/s41558-018-0180-3

Haddaway, N. R., Macura, B., Whaley, P., & 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*, 7. https://doi.org/10.1186/s13750-018-0121-7

Haddaway, N. R., & Verhoeven, J. T. A. (2015). Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecology and Evolution*, *5*, 4451–4454. https://doi.org/10.1002/ece3.1722

Hartling, L. (2009). Risk of Bias versus quality assessment of randomised controlled trials: Cross sectional study. *BMJ*, *339*, b4012. https://doi.org/10.1136/bmj.b4012

Hartling, L., Hamm, M. P., Milnea, A., Vandermeer, B., Santaguid, P. L., Ansari, M., Tsertsvadze, A., Hempel, S., Shekelle, P., & Dryden, D. M. (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*, 973–981. https://doi.org/10.1016/j.jclinepi.2012.07.005

Hennessy, E. A., Acabchuk, R. L., Arnold, P. A., Dunn, A. G., Foo, Y. Z., Johnson, B. T., Geange, S. R., Haddaway, N. R., Nakagawa, S., Mapanga, W., Mengersen, K., Page, M. J., Sánchez-Tójar, A., Welch, V., & McGuinness, L. A. (2022). Ensuring prevention science research is synthesis-ready for immediate and lasting scientific impact. *Prevention Science*, *23*, 809–820. https://doi.org/10.1007/s11121-021-01279-8

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, *15*, 615–625. http://www.jstor.org/stable/20485961

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., Sterne, J. A. C., Cochrane Bias Methods Group, & Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*, d5928. https://doi.org/10.1136/bmj.d5928

Hillebrand, H., & Gurevitch, J. (2013). Reporting standards in experimental studies. *Ecology Letters*, *16*, 1419–1420. https://doi.org/10.1111/ele.12190

Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLoS Biology*, *13*, e1002190. https://doi.org/10.1371/journal.pbio.1002190

Hooijmans, C. R., Rovers, M. M., de Vries, R. B., Leenaars, M., Ritskes-Hoitinga, M., & Langendam, M. W. (2014). SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology*, *14*, 43. https://doi.org/10.1186/1471-2288-14-43

Ihle, M., Winney, I. S., Krystalli, A., & Croucher, M. (2017). Striving for transparent and credible research: Practical guidelines for behavioral ecologists. *Behavioral Ecology*, *28*, 348–354. https://doi.org/10.1093/beheco/arx003

Ivimey-Cook, E. R., Sánchez-Tójar, A., Berberi, I., Culina, A., Roche, D. G., Almeida, R. A., Amin, B., Bairos-Novak, K. R., Balti, H., Bertram, M. G., Bliard, L., Byrne, I., Chan, Y., Cioffi, W. R., Corbel, Q., Elsy, A. D., Florko, K. R. N., Gould, E., Grainger, M. J., … Moran, N. P. (2025). From policy to practice: Progress towards data- and code-sharing in ecology and evolution. *Proceedings of the Biological Sciences*, *292*, 20251394. https://doi.org/10.1098/rspb.2025.1394

Jefferson, T., Jones, M. A., Doshi, P., Del Mar, C. B., Hama, R., Thompson, M. J., Spencer, E. A., Onakpoya, I., Mahtani, K. R., Nunan, D., Howick, J., & Heneghan, C. J. (2014). Neuraminidase inhibitors for preventing and treating influenza in adults and children. *Cochrane Database of Systematic Reviews*, *10*, CD008965. https://doi.org/10.1002/14651858

Jeyaraman, M. M., Rabbani, R., Copstein, L., Robsonc, C. R., Al-Yousif, N., Pollock, M., Xia, J., Balijepalli, C., Hofer, K., Mansour, S., Fazeli, M. S., Ansari, M. T., Tricco, A. C., & Abou-Setta, A. M. (2020). Methodologically rigorous risk of bias tools for nonrandomized studies had low reliability and high evaluator burden. *Journal of Clinical Epidemiology*, *128*, 140–147. 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

Kaiser, L., Wat, C., Mills, T., Mahoney, P., Ward, P., & Hayden, F. (2003). Impact of Oseltamivir treatment on influenza-related lower respiratory tract complications and hospitalizations. *Archives of Internal Medicine*, *163*, 1667–1672. https://doi.org/10.1001/archinte.163.14.1667

Kardish, M. R., Mueller, U. G., Amador-Vargas, S., Dietrich, E. I., Ma, R., Barrett, B., & Fang, C.-C. (2015). Blind Trust in Unblinded Observation in ecology, evolution, and behavior. *Frontiers in Ecology and Evolution*, *3*, 51. https://doi.org/10.3389/fevo.2015.00051

Kim, B., Moran, N. P., Reinhold, K., & Sánchez-Tójar, A. (2021). Male size and reproductive performance in three species of livebearing fishes (*Gambusia* spp.): A systematic review and meta-analysis. *Journal of Animal Ecology*, *90*, 2431–2445. https://doi.org/10.1111/1365-2656.13554

Konno, K., Akasaka, M., Koshida, C., Katayama, N., Osada, N., Spake, R., & Amano, T. (2020). Ignoring non-English-language studies may bias ecological meta-analyses. *Ecology and Evolution*, *10*, 6373–6384. https://doi.org/10.1002/ece3.6368

Konno, K., Gibbons, J., Lewis, R., & Pullin, A. S. (2025). Risk of bias in effect estimates is real but still largely ignored in environmental research. *Next Research*, *2*, 100565. https://doi.org/10.1016/j.nexres.2025.100565

Konno, K., Livoreil, B., & 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/

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

Kozlov, M. V., & Zvereva, E. L. (2015). Confirmation bias in studies of fluctuating asymmetry. *Ecological Indicators*, *57*, 293–297. https://doi.org/10.1016/j.ecolind.2015.05.014

Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying quantitative bias analysis to epidemiologic data*. Springer Science & Business Media.

Lodi, S., Godoy, B. S., Ortega, J. C. G., & Bini, L. M. (2021). Quality of meta-analyses in freshwater ecology: A systematic review. *Freshwater Biology*, *66*, 803–814. https://doi.org/10.1111/fwb.13695

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, *526*, 187–189. https://doi.org/10.1038/526187a

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

Martin, L. J., Blossey, B., & 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., & Sánchez-Tójar, A. (2025). Do egg hormones have fitness consequences in wild birds? A systematic review and meta-analysis. *Ecology Letters*, *28*, e70100. https://doi.org/10.1111/ele.70100

Minozzi, S., Cinquini, M., Gianola, S., Castellini, G., Gerardi, C., & 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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097. https://doi.org/10.1371/journal.pmed.1000097

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

Mukherjee, N., Hugé, J., Sutherland, W. J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., & Koedam, N. (2015). The Delphi technique in ecology and biological conservation: Applications and guidelines. *Methods in Ecology and Evolution*, *6*, 1097–1109. https://doi.org/10.1111/2041-210X.12387

Nakagawa, S., Dunn, A. G., Lagisz, M., Bannach-Brown, A., Grames, E. M., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Westgate, M. J., Arnold, P. A., Barrow, S., Bethel, A., Cooper, E., Foo, Y. Z., Geange, S. R., Hennessy, E., Mapanga, W., Mengersen, K., Munera, C., … Haddaway, N. R. (2020). A new ecosystem for evidence synthesis. *Nature Ecology & Evolution*, *4*, 498–501. https://doi.org/10.1038/s41559-020-1153-2

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*, 4–21. https://doi.org/10.1111/2041-210X.13724

Nuñez, M. A., & Amano, T. (2021). Monolingual searches can limit and bias results in global literature reviews. *Nature Ecology & Evolution*, *5*, 264. https://doi.org/10.1038/s41559-020-01369-w

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., & Moher, D. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biological Reviews*, *96*, 1695–1722.

Page, M. J., & Higgins, J. P. T. (2016). Rethinking the assessment of risk of bias due to selective reporting: A cross-sectional study. *Systematic Reviews*, *5*, 108. https://doi.org/10.1186/s13643-016-0289-2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., En Chee, Y., Kelly, C. D., Gurevitch, J., & Nakagawa, S. (2016). Transparency in ecology and evolution: Real problems, real solutions. *Trends in Ecology & Evolution*, *31*, 711–719. https://doi.org/10.1016/j.tree.2016.07.002

Pescott, O. L., Boyd, R. B., Powney, G. D., & 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

Petersen, J. M., Barrett, M., Ahrens, K. A., Murray, E. J., Bryant, A. S., Hogue, C. J., Mumford, S. L., Gadupudi, S., Fox, M. P., & Trinquart, L. (2022). The confounder matrix: A tool to assess confounding bias in systematic reviews of observational studies of etiology. *Research Synthesis Methods*, *13*, 242–254. https://doi.org/10.1002/jrsm.1544

Pildal, J., Hróbjartsson, A., Jørgensen, K. J., Hilden, J., Altman, D. G., & Gøtzsche, P. C. (2007). Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *International Journal of Epidemiology*, *4*, 847–857. https://doi.org/10.1093/ije/dym087

Purgar, M., Klanjscek, T., & Culina, A. (2022). Quantifying research waste in ecology. *Nature Ecology & Evolution*, *6*, 1390–1397. 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. https://www.R-project.org/

Sandoval-Lentisco, A., López-López, J. A., & 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*, 1244–1245. 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*, 89–104. https://doi.org/10.1093/evlett/qrae051

Sedgwick, P. (2014). Randomised controlled trials: Internal versus external validity. *BMJ*, *348*, 1742. https://doi.org/10.1136/bmj.g1742

Senn, S. (2013). Seven myths of randomisation in clinical trials. *Statistics in Medicine*, *32*, 1439–1450.

Stanhope, J., & Weinstein, P. (2023). Critical appraisal in ecology: What tools are available, and what is being used in systematic reviews? *Research Synthesis Methods*, *14*, 342–356. https://doi.org/10.1002/jrsm.1609

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Sterne, J. A. C., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., … Higgins, J. P. T. (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., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*, 9132. https://doi.org/10.1038/s41598-017-09084-6

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

Wang, Y., Ghadimi, M., Wang, Q., Hou, L., Zeraatkar, D., Iqbal, A., Ho, C., Yao, L., Hu, M., Ye, Z., Couban, R., Armijo-Olivo, S., Bassler, D., Briel, M., Gluud, L. L., Glasziou, P., Jackson, R., Keitz, S. A., Letelier, L. M., … Guyatt, G. H. (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*, 218–225. https://doi.org/10.1016/j.jclinepi.2022.10.018

Westgate, M. J. (2019a). Revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods*, *10*, 606–614. https://doi.org/10.1002/jrsm.1374

Westgate, M. J. (2019b). *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*, 292–298. https://doi.org/10.1093/aje/kws412

Wickham, H. (2023). *"stringr"*. https://doi.org/10.32614/CRAN.package.stringr

Yang, Y., Noble, D. W. A., Spake, R., Senior, A. M., Lagisz, M., & Nakagawa, S. (2025). A pluralistic framework for measuring, interpreting and decomposing heterogeneity in meta-analysis. *Methods in Ecology and Evolution*, *16*, 2710–2725. https://doi.org/10.1111/2041-210x.70155

Methods in Ecology and Evolution

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1. (1) Survey on the familiarity with the risk of bias, (2) Email text/ call to participate in the survey, (3) Search string for identifying meta-analyses published across 300 journals and (4) Journals and the Risk of Bias - data extraction form.

How to cite this article: Culina, A., Foster, D., Grainger, M., O'Dea, R. E., Pescott, O. L., Sánchez-Tójar, A., Boyd, R. J., Koricheva, J., Nakagawa, S., & Stewart, G. (2026). Elevating the importance of Risk of Bias assessment for ecology and evolutionary biology. *Methods in Ecology and Evolution*, *00*, 1–15. https://doi.org/10.1111/2041-210x.70356