ROB-ME: a tool for assessing risk of bias due to missing evidence in systematic reviews with meta-analysis

Matthew J Page, senior research fellow[1,*], Jonathan AC Sterne, professor[2], Isabelle Boutron, professor[3], Asbjørn Hróbjartsson, professor[4], Jamie J Kirkham, professor[5], Tianjing Li, associate professor[6], Andreas Lundh, associate professor[7], Evan Mayo-Wilson, associate professor[8], Joanne E McKenzie, professor[1], Lesley A Stewart, professor[9], Alex J Sutton, professor[10], Lisa Bero, professor[11], Adam G Dunn, professor[12], Kerry Dwan, senior lecturer[13], Roy G Elbers, senior researcher[14], Raju Kanukula, doctoral student[1], Joerg J Meerpohl, professor[15], Erick H Turner, associate professor[16], Julian PT Higgins, professor[17]

1. Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; National Institute for Health and Care Research Bristol Biomedical Research Centre, Bristol, UK; Health Data Research UK South-West, Bristol, UK
3. Université de Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, F 75004 Paris, France
4. Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark; Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark
5. Centre for Biostatistics, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
6. Department of Ophthalmology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA
7. Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark; Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark; Department of Respiratory Medicine and Infectious Diseases, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Denmark
8. Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, North Carolina, United States
9. Centre for Reviews and Dissemination, University of York, York, UK
10. Department of Population Health Sciences, University of Leicester, Leicester, UK
11. Center for Bioethics and Humanities, University of Colorado, Colorado, USA

12. Biomedical Informatics and Digital Health, School of Medical Sciences, The University of Sydney, Sydney, Australia
13. Liverpool School of Tropical Medicine, Liverpool, UK
14. Department of General Practice, Intellectual Disability Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
15. Institute for Evidence in Medicine, Medical Center & Faculty of Medicine, University of Freiburg, Freiburg, Germany; Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany
16. Department of Psychiatry, Oregon Health & Science University, Portland, Oregon, USA; Behavioral Health and Neurosciences Division, Veterans Affairs Portland Health Care System, Portland, Oregon, USA
17. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; National Institute for Health and Care Research Bristol Biomedical Research Centre, Bristol, UK

Correspondence to: Dr. Matthew Page, Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, 3004, Australia. Telephone: +61 9903 0248. Email address: matthew.page@monash.edu

Word count: 4,013

STANDFIRST

Various methods are available to help users assess whether selective non-publication of studies or selective non-reporting of study results has occurred, but not its impact on a meta-analysis. This leaves users to decide their own approach for judging the risk of bias in a meta-analysis result. In this paper, we describe the ROB-ME (Risk Of Bias due to Missing Evidence) tool, the first structured approach for assessing the risk of bias that arises when entire studies, or particular results within studies, are missing from a meta-analysis because of the P value, magnitude or direction of the study results. We anticipate that the tool will help authors and users of systematic reviews identify meta-analyses at high risk of bias and interpret results appropriately.

SUMMARY POINTS

- ~~Researchers' decisions about whether to r~~The reporting of primary studies or results ~~are often~~might be influenced by the P value, magnitude or direction of the study results. This can lead to bias in a meta-analysis, because the available evidence (from studies or results) differs systematically from the missing evidence.
- Existing methods mostly help users assess whether selective non-publication of studies or selective non-reporting of study results has occurred, but not its impact on a meta-analysis. This leaves users to decide their own approach for combining the risks of each type of bias into an overall judgement of risk of bias in a meta-analysis result.
- The ROB-ME (Risk Of Bias due to Missing Evidence) tool is the first structured approach for assessing the risk of bias that arises when entire studies, or particular results within studies, are missing from a meta-analysis because of the P value, magnitude or direction of the study results. The tool consists of three preliminary steps – select and define which meta-analyses will be assessed; determine which studies meeting the inclusion criteria for the meta-analyses have missing results; and consider the potential for missing studies across the review – which inform an assessment of risk of bias due to missing evidence in a particular meta-analysis result.
- We anticipate the ROB-ME tool will help authors and users of systematic reviews identify meta-analyses at high risk of bias and interpret results appropriately.

INTRODUCTION

A key feature of systematic reviews of quantitative research is the attempt to identify all studies that meet the review inclusion criteria and to include relevant data from all such studies in meta-analyses. This goal is compromised when reporting ~~decisions are~~of primary studies is influenced by the P value, magnitude or direction of study results (1). These factors ~~may inform the decision not to publish a study report~~might influence whether a study is published at all ('selective non-publication of studies' or 'publication bias') (2, 3), ~~to delay publication of~~the speed at which a study report is published ('time-lag bias') (4) or ~~to disseminate the report in a non-indexed~~type of journal (indexed or not) in which a study report is published ('location bias') (5), each of which can lead to *studies* missing from meta-analyses. The P value, magnitude or direction of the study results ~~may also inform decisions not to report~~might also influence whether, or how completely, particular results ~~or to~~are reported ~~them incompletely~~ ('selective non-reporting of study results' or 'outcome reporting bias') (6), leading to *results* missing from meta-analyses even when the study has been identified. The term 'reporting bias' has often been used to describe such selective dissemination of evidence, but here we use the term 'non-reporting bias' to emphasize the non-availability of evidence (7).

We present some examples of non-reporting bias to explain the concepts above. Suppose that after conducting a randomized trial comparing glucocorticoid injection with placebo for shoulder pain, investigators find that shoulder strength was greater in participants receiving glucocorticoid injection, but pain intensity and function were no different between groups. The investigators might never submit a study report for publication, because the observed results conflict with their prior hypotheses about the benefits of glucocorticoid injection, or because they assume journal editors will not be interested in publishing the paper. This is an example of selective non-publication of studies ('publication bias'). Alternatively, suppose that a study report is published but results for pain intensity and function are omitted entirely or presented incompletely; for example, a statement that pain and function scores were "not different between groups", without summary statistics, effect estimates or measures of precision. This is an example of selective non-reporting of study results ('outcome reporting bias').

A meta-analysis result will be biased when the available evidence (from studies or results) differs systematically from the missing evidence. However, existing methods have limited ability to determine which meta-analyses are at risk of bias due to missing evidence (8, 9). For example, several tools have been developed to help users assess whether selective non-reporting of study results has occurred, but not its impact on a meta-analysis. Furthermore, current methods tend to focus on only

one type of non-reporting bias (only selective non-publication of studies, or only selective non-reporting of study results). This leaves users to decide their own approach for combining the risks of each type of bias into an overall judgement of risk of bias in the meta-analysis result; likely leading to inconsistency in judgements. In this paper, we describe the ROB-ME (Risk Of Bias due to Missing Evidence) tool, the first structured approach for assessing the risk of bias that arises when entire studies, or particular results within studies, are missing from a meta-analysis because of the P value, magnitude or direction of the study results.

DEVELOPMENT OF THE ROB-ME TOOL

We followed the framework for developing risk of bias tools recommended by Whiting et al. (10). A core group (MJP, JACS and JPTH) coordinated development of the tool, which included assembling the team of collaborators, preparing meeting materials, and leading the drafting and revising of the tool. Preliminary work to inform development of the tool included a cross-sectional study of how selective non-reporting of study results was assessed in Cochrane reviews (11), a systematic review of scales, checklists and domain-based tools for assessing risk of non-reporting biases in studies and meta-analyses of studies (8), and a non-systematic review of the empirical evidence of non-reporting biases (5).

Informed by the preliminary work, the core group developed an initial proposal for a new tool for assessing risk of bias due to missing evidence in a meta-analysis result and presented it at a development meeting in April 2017. Seventeen contributors with expertise in the empirical evidence of non-reporting biases, graphical and statistical approaches to assess non-reporting biases, and methods for identifying and accessing trial protocols, trials register entries, and information submitted to regulators (e.g. clinical study reports) attended the meeting. Through a series of presentations and facilitated discussion sessions, meeting participants agreed on the scope, structure and content to be addressed by the new tool and identified topics for further consideration.

Following the development meeting, the lead author (MJP) prepared initial drafts of the ROB-ME tool and discussed them with meeting participants via videoconferences in August and November 2017. The core group developed the tool, the assessment framework underpinning it and accompanying guidance further between 2018 and 2020 while drafting a chapter on the topic for the 2019 edition of the *Cochrane Handbook for Systematic Reviews of Interventions* (7). Additional edits were made in response to feedback received on a draft of the tool presented at the 2018 Cochrane Colloquium (12) and a draft sent to all co-authors of this paper. A preliminary version of the tool (template and detailed

guidance) was uploaded to https://www.riskofbias.info/ and presented in a webinar in October 2020, at which systematic reviewers were invited to provide feedback via a template form seeking views on each component of the tool. Twelve systematic reviewers subsequently provided written feedback via the template form or by commenting directly on the tool template; six had piloted the tool on a meta-analysis they were conducting prior to providing feedback. All systematic reviewers viewed the tool favourably, although some edits to the instructions and wording of questions (but not to the structure of the tool) were suggested. This written feedback, along with verbal feedback received from attendees of seven webinars delivered throughout 2021 and 2022 was considered by the core team, who revised the tool by amending the wording to improve clarity. The final version was sent to all co-authors for approval.

THE ROB-ME TOOL

The full ROB-ME tool is available at https://www.riskofbias.info/. Four worked examples of applying the ROB-ME tool are provided in Appendix 1-4 of the Supplement.

Terminology

Throughout this document, we use the phrase 'study outcome' to refer to an outcome measurement collected on, or by, participants in a study; a measurement could be continuous or non-continuous (e.g. a binary event or rate). We use the phrase 'study result' to describe the combination of a point estimate and a measure of its precision (or the summary statistics required to calculate these) for a particular study outcome (7, 13). An example of a study outcome is "pain at the end of eight weeks of treatment, measured using a 100-point visual analogue scale". A corresponding study result for this outcome might be an estimated difference in mean pain scores between intervention groups with 95% confidence interval.

Scope of the tool

The ROB-ME tool is designed for authors or users of systematic reviews to assess risk of bias due to missing evidence in a pairwise meta-analysis of the effect of interventions. It can be applied regardless of the number and types of studies with results available for inclusion in the meta-analysis, including in cases where only one of the studies identified has results available. The tool is not designed for assessing risk of bias due to missing evidence in a network meta-analysis; a tool for this purpose (ROB-MEN) is available and described elsewhere (14).

The ROB-ME tool is not intended to examine a related source of bias, which we call 'risk of bias in selection of the reported result', which arises when an *available* study result was selected for reporting by the study authors from among multiple measurements or analyses, on the basis of the P value, magnitude or direction of these multiple results (15). For example, study investigators might measure pain using two measurement instruments, yet report results only for the instrument that yielded a statistically significant effect estimate. In this circumstance, a study result is available for inclusion in a meta-analysis of pain scores, although the study result is at high risk of bias because of the way it was selected for reporting. Risk of bias in selection of the reported result is considered in tools designed to assess risk of bias in a *study result* (such as the RoB 2 tool for assessing risk of bias in randomized trials (15) and the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions (16)). By contrast, ROB-ME is used to assess risk of bias in a *meta-analysis result* and addresses the risk of bias arising from omission of results from one or more studies from the meta-analysis. ROB-ME is therefore not designed to replace the assessment of risk of bias in selection of the reported result in a study, and so users assessing that source of bias should continue to apply RoB 2 or ROBINS-I as appropriate, in addition to ROB-ME. Table 1~~Table 1~~ specifies which risk-of-bias tool to use when confronted with different reporting scenarios.

How to conduct ROB-ME assessments

Application of the ROB-ME tool to a systematic review consists of four steps (Figure 1~~Figure 1~~):

1. Select and define which meta-analyses will be assessed for risk of bias due to missing evidence.
2. Determine which studies meeting the inclusion criteria for the meta-analyses have missing results and thus cannot contribute to the meta-analyses.
3. Consider the potential for missing studies across the systematic review.
4. Assess risk of bias due to missing evidence in each meta-analysis.

We recommend that the tool be completed by ~~multiple~~ at least two users independently, with any discrepancies resolved via discussion, or adjudication from another user. We recommend that Step 1 be conducted at the protocol stage, Step 2 during data collection or when assessing risk of bias in the results of included studies, and Step 3 and 4 after generating meta-analysis results, but before assessing certainty in the body of evidence (e.g. via GRADE (17)).

Planning the risk-of-bias assessment

Assessing risk of bias due to missing evidence in all meta-analyses in a systematic review might not be possible (e.g. when resources are limited). In such cases, users should select which meta-analyses to

assess for risk of bias based on which outcomes are most important for decision making (Step 1). Such outcomes tend to be those selected for inclusion in 'Summary of findings' tables (18). Ideally, users should pre-specify the meta-analyses they intend to assess for risk of bias and indicate the 'PICO' (Population, Intervention, Comparator, Outcome) for each meta-analysis (that is, define the question that each meta-analysis aims to answer) (7, 19). Users should also seek to define fully the types of studies and results that are eligible for inclusion in each meta-analysis to be assessed for risk of bias: doing so helps clarify which results are missing.

Evidence of selective non-reporting of results should then be gathered for each study that is eligible for inclusion in the meta-analyses selected for ROB-ME assessments (Step 2). Users should start by assembling available sources of information about each eligible study. This might include the trials register entry (e.g. at ClinicalTrials.gov), study protocol, statistical analysis plan, reports of results of the study (e.g. journal article, clinical study report), or information obtained directly from the study authors or sponsor (e.g. data files supplied). ~~Building on the ORBIT (Outcome Reporting Bias In Trials) approach (6), we recommend that users of ROB-ME record in a matrix whether results were available for each study (see examples of completed matrices in the Supplement).~~ If study plans are available, for example in a trials register entry (20), study protocol, or statistical analysis plan, then details of pre-specified outcomes ~~can~~ should be compared with other sources of information about a study, such as a journal article presenting the main findings, to identify any outcomes without results available. Users might find it helpful to construct a matrix for each eligible study which lists all outcomes described in the study plans and records whether results were available for each. If no study plans are available, users can cross-check the methods and results sections against one another to identify any outcomes with no results reported, or results reported incompletely (i.e. without both a point estimate and measure of precision, or means of deriving these).

Once users have identified that a study does not have a result available for inclusion in a meta-analysis, they should consider why the result is unavailable. Possible reasons include: the outcome was not measured at all by the study investigators; the outcome was measured but not analysed by study investigators (for example, due to a substantial amount of missing data); the outcome was measured and analysed but the result did not support the investigators' hypothesis (e.g. a statistically non-significant result was observed in a superiority trial); or another reason, which may or may not be related to the nature of the result (see Box 1~~Box 1~~). Building on the ORBIT (Outcome Reporting Bias In Trials) approach (6), we recommend that users of ROB-ME record in a matrix whether results were available for each study meeting the inclusion criteria for the meta-analyses, which lists each study in

rows and each meta-analysis to be assessed for risk of bias in columns (see examples of completed matrices in the Supplement). When completing the matrix, assessments of availability of results should be based on all the sources of information about a study obtained by users. That is, if harms results were not reported in a journal article but were provided by study authors upon request, then ROB-ME users should specify in the matrix that they have a study result available for inclusion in the synthesis.

Users of ROB-ME should then consider whether it is possible that some eligible studies, in addition to those considered in Step 2, were not identified because of the P value, magnitude or direction of their results (Step 3). We anticipate this will be the case for most systematic reviews, apart from those for which an inception cohort was defined (e.g. only prospectively registered studies or studies identified for a prospective meta-analysis were eligible for inclusion in the review). In such reviews, all such studies are identified before their results became known.

Assessing a specific meta-analysis result

In Step 4, users of the ROB-ME tool answer eight signalling questions, which seek to elicit information about what happened or was observed (Table 2~~Table 2~~). Answers to the signalling questions are informed by the material collated in Steps 1 to 3. Step 4 should be completed for each meta-analysis assessed for risk of bias.

The response options for the signalling questions are: Yes (Y); Probably yes (PY); Probably no (PN); No (N); No information (NI); Not applicable (NA). To maximize simplicity and clarity, the questions are phrased such that a response of 'Yes' indicates higher risk of bias and 'No' indicates lower risk of bias. Responses of 'Yes' and 'Probably yes' have the same implications for risk of bias, as do responses of 'No' and 'Probably no'. The definitive versions ('Yes' and 'No') would typically be selected when firm evidence is available in relation to the signalling question, whereas the 'Probably' versions would typically be selected when firm evidence is lacking and some judgement has been made. Guidance on how to answer each signalling question is provided in the tool available at https://www.riskofbias.info/.

Signalling questions relating to the within-study assessment of non-reporting bias

The first four signalling questions ask users to consider the extent of missing results in the studies identified for the meta-analysis being assessed for risk of bias ('known unknowns'). If they determine that one or more of the studies is missing from the meta-analysis because of selective non-reporting

of study results, users should consider whether the amount of missing evidence matters: i.e. whether inclusion of the omitted results would likely lead to a notable change in the summary (combined) effect estimate, given the likely weight and direction of effect in studies omitted from the meta-analysis (if known), meta-analysis model used (such as fixed-effect or random-effects) and extent of heterogeneity observed. To clarify the potential for bias in the meta-analysis result, users could generate a forest plot displaying studies with results, along with information about studies known to be missing from the meta-analysis due to selective non-reporting of their results (see ~~Figure 1~~ Figure 2~~Figure 2~~ for an example).

Signalling questions relating to the across-study assessment of non-reporting bias

The remaining four signalling questions ask users to consider the risk that the meta-analysis result is biased because additional studies or study results, beyond those already identified, are missing ('unknown unknowns'). Factors to consider include whether missing studies are likely to have had eligible results because the outcome is typically measured in all studies on the topic; whether the pattern of results included in the meta-analysis reveals a tendency for studies with particular results, such as those with $P > 0.05$, to be missing (as observed through graphical methods such as contour-enhanced funnel plots (21)); and whether the findings of appropriate sensitivity analyses, which might include restricting the meta-analysis to the largest studies or using selection models (22, 23, 24) or regression-based adjustment methods (25, 26), suggest that a meta-analysis result is not robust to plausible assumptions about the extent and nature of missing evidence.

Risk-of-bias judgement

ROB-ME operates in the same manner as the RoB 2 (15) and ROBINS-I (16) tools, in which responses to signalling questions provide the basis for a judgement about the risk of bias in the specific meta-analysis result being assessed. The tool includes an algorithm that maps responses to signalling questions onto a proposed risk-of-bias judgement (see Appendix 5 and 6 in the Supplement). The possible risk-of-bias judgements are:

1. Low risk of bias: the meta-analysis result is *unlikely* to be biased due to missing evidence;
2. High risk of bias: the meta-analysis result is *likely* to be biased due to missing evidence;
3. Some concerns: there are uncertainties about the extent or potential impact of missing evidence that preclude a judgement of low or high risk of bias.

Although ROB-ME considers only the summary effect estimate, we recognise that suppression of results may impact other statistics, such as the estimate of heterogeneity, and in turn the width of the confidence interval for the summary effect estimate.

Presentations of risk-of-bias assessments

Users of ROB-ME should present risk-of-bias judgements in the main systematic review report (for example in a table or within the forest plot), along with a brief free-text justification for each judgement. In addition, we encourage reporting of the completed results matrix (Step 2) and answers to all questions in Steps 3 and 4 (with supporting text, where applicable) as supplementary material. Only consensus judgements and answers, rather than judgements and answers from individual users, should be presented.

DISCUSSION

Inadequate consideration of risk of bias due to missing evidence could lead to ineffective or harmful treatments being recommended but, despite the implementation of various initiatives to address the problem, selective non-publication of studies and selective non-reporting of study results persists (2, 3). For this reason, systematic reviewers should routinely assess the possibility that these issues have biased the results of meta-analyses they have conducted. We developed the ROB-ME tool to help reviewers undertake these assessments. The risk-of-bias judgements drawn from ROB-ME should help distinguish stronger from weaker synthesised evidence and influence the certainty of conclusions drawn from a systematic review (potentially as part of a GRADE assessment (17)).

The ROB-ME tool includes several innovations in the assessment of non-reporting biases. In the original Cochrane tool for assessing risk of bias in randomized trials (27), users were prompted to judge the risk of selective reporting bias at the study-level, based on whether *any* results in the study were selectively (non-)reported. In reviews adopting this approach, many *studies* have been judged at high risk of selective reporting bias (11); however, the corresponding risk of bias in *meta-analyses* affected by selective non-reporting of study results is infrequently acknowledged, because no guidance on how to reach such a judgement was provided. ROB-ME explicitly addresses this gap, directing assessments at the level of the meta-analysis result and outlining what factors need to be considered to determine whether the amount of evidence known or assumed to be missing matters. Furthermore, ROB-ME is the first tool to help users reach an overall judgement about risk of bias in a meta-analysis result arising from both missing studies and missing results in studies.

Of the various components of ROB-ME, we anticipate that within-study assessment of non-reporting bias will be the most resource intensive, yet also the most valuable. This is because the impact of selective non-reporting of results in a set of studies known to be missing from a meta-analysis can be

quantified more easily than the impact of selective non-publication of an unknown number of studies. Furthermore, if systematic reviewers suspect that a meta-analysis result is biased because results were missing selectively from a large proportion of the studies identified, then the across-study assessment of non-reporting bias is unlikely to change their judgement (other than increasing their certainty that the meta-analysis result is at high risk of bias). This does not imply that the across-study assessment should be considered an optional component of the tool. Despite their well-known limitations (1, 28), methods originally developed for an across-study assessment (such as funnel plots, tests for funnel plot asymmetry and sensitivity analyses) are useful when the within-study assessment of selective non-reporting is limited (e.g. when detailed study plans are unavailable for most studies included in the review).

The ROB-ME tool was designed to complement tools such as RoB 2 (15) and ROBINS-I (16) for assessing risk of bias in study results. These tools enable assessment of bias in selection of the reported result, a domain of bias related to, but not addressed by, ROB-ME. As a rule of thumb, ROB-ME should be used when systematic reviewers do not have a result from a study to include in a particular meta-analysis, whereas RoB 2 or ROBINS-I should be used to assess whether a result that is available for inclusion might have been "cherry-picked" (29) from among multiple measures or analyses. Assessments of risk of bias due to missing evidence and risk of bias in selection of the reported result will likely be informed by the same sources of information about studies (e.g. study protocols, register entries), so we advise systematic reviewers to apply Step 2 of ROB-ME (completion of the results matrix) in parallel with RoB 2 or ROBINS-I.

The ROB-ME tool is most suitable for assessing meta-analyses of evidence from randomized trials. We believe it can also be used to assess meta-analyses of non-randomized studies of interventions (e.g. cohort studies, interrupted time series studies), but some components of the tool will not apply to such studies. For example, the applicability of tests for funnel plot asymmetry in the context of meta-analyses of non-randomized studies of interventions is unclear (28), so these should not be used in the across-study assessment of non-reporting bias. Furthermore, analyses of non-randomized studies are less likely than randomized trials to be registered or have a publicly accessible protocol, so comparison of pre-specified with reported outcomes will usually not be possible. However, assessment of selective non-reporting of study results can still be undertaken for such studies, for example by comparing the methods and results sections of a study report.

Plans for further development of the ROB-ME tool include further refining to ensure it is suitable for assessing types of pairwise synthesis other than meta-analysis that yield a point estimate of an intervention effect (such as calculation of the median effect across studies when meta-analysis is not possible or appropriate) (30, 31). We also plan to develop an interactive online version of the tool to facilitate use and prepare a bank of worked examples for educational purposes. We will also explore how ROB-ME judgements should feed into the GRADE framework for assessing certainty in the body of evidence. We welcome feedback from users of ROB-ME and any subsequent updates to the tool will be uploaded to https://www.riskofbias.info/.

CONCLUSION

The ROB-ME tool addresses gaps in existing approaches for assessing risk of non-reporting biases. We hope that the tool will be useful to authors and users of systematic reviews, by helping to identify meta-analyses at high risk of bias and facilitating appropriate interpretation of results.

DECLARATIONS

Acknowledgments

We thank Sarah Dawson, Kay Dickersin, Carl Heneghan, Toby Lasserson, Hannah Rothstein and James Thomas for their participation at the development meeting and comments on early drafts of the ROB-ME tool. We thank the following contributors who provided feedback on a preliminary version of the tool: Robert Boyle, Mahnuma Estee, Alix Hall, Kylie Hunter, Alison Su-Hsun Liu, David Marshall, Paul McCann, Riaz Qureshi, Jelena Savović, Rik van Eekelen, Rui Wang, and Rebecca Wilson. We thank Phi-Yen Nguyen for providing feedback on a more advanced version of the tool.

Contributions

All authors declare to meet the ICMJE conditions for authorship. MJP, JACS and JPTH conceived the project. MJP, JACS and JPTH oversaw the project. MJP led the research to inform the tool. MJP prepared all materials for the development meeting. MJP, JACS, IB, AH, TL, LAS, AJS, SD, KD, JJM, EHT and JPTH attended and contributed to the development meeting. RGE took notes at the development meeting. MJP, JACS and JPTH wrote the first draft of the ROB-ME tool. All authors contributed to the development of the ROB-ME tool and to writing associated guidance. MJP led the drafting of this manuscript. All authors reviewed and commented on drafts of the manuscript. All authors approved the final version of the manuscript. MJP is the guarantor of this work.

Funding

MJP was supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535) and an Australian Research Council Discovery Early Career Researcher Award (DE200101618) during the conduct of this research. JACS is supported by the NIHR Bristol Biomedical Research Centre and Health Data Research UK South-West and is a member of the MRC Integrative Epidemiology Unit at the University of Bristol. TL was supported by grant UG1 EY020522 from the National Eye Institute, National Institutes of Health. JEM is supported by an Australian NHMRC Investigator Grant (GNT2009612). JPTH is a National Institute for Health and Care Research (NIHR) Senior Investigator (NF-SI-0617-10145), is supported by the NIHR Bristol Biomedical Research Centre and the NIHR Applied Research Collaboration West (ARC West) at University Hospitals Bristol NHS Foundation Trust and is a member of the MRC Integrative Epidemiology Unit at the University of Bristol. Development of the ROB-ME tool, writing the paper and the decision to submit for publication were independent of all research funders.

Competing interests

All authors have completed the ICMJE disclosure form at http://www.icmje.org/disclosure-of-interest/ and declare: JJK is a statistical editor for the BMJ, for which he receives remuneration. EMW has received grants/contracts from the Patient Centered Outcomes Research Institute, National Institutes of Health, Arnold Ventures and Robert Wood Johnson Foundation, has received consulting fees for editorial work from Origin Editorial, payment/honoraria from the American Public Health Association and National Institutes of Health and has an unpaid leadership role in the American Psychological Association. LB has received grants/contracts from the State of Colorado, Cochrane, and the US Environmental Protection Agency, has received consulting fees for work as a conflict of interest advisor from Health Canada, has received support for travel from the National Academy of Science Engineering and Medicine, Cochrane and the World Congress on Research Integrity, and has an unpaid Chair role for the National Academy of Science Engineering and Medicine committees. ET has received an honorarium for a lecture on publication bias from the University of British Columbia, Canada. All other authors declare having no competing interests.

Patient and Public Involvement

Patients and members of the public were not involved in this methodological research. Our motivation for developing the ROB-ME tool arose from our concerns as people who interact with the healthcare system that bias due to missing evidence in meta-analyses can potentially lead to ineffective or harmful treatments being delivered to patients. We plan to disseminate the research widely, including

to community participants in evidence synthesis organisations, as we believe increased awareness about non-reporting biases and its consequences can help minimise the problem.

Provenance and peer review

Not commissioned; externally peer reviewed.

REFERENCES

1. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Research Synthesis Methods. 2021;12(2):248-59.
2. Turner EH, Cipriani A, Furukawa TA, Salanti G, de Vries YA. Selective publication of antidepressant trials and its influence on apparent efficacy: Updated comparisons and meta-analyses of newer versus older trials. PLoS Medicine. 2022;19(1):e1003886.
3. Speich B, Gryaznov D, Busse JW, Gloy VL, Lohner S, Klatte K, et al. Nonregistration, discontinuation, and nonpublication of randomized trials: A repeated metaresearch analysis. PLoS Medicine. 2022;19(4):e1003980.
4. Song SY, Koo DH, Jung SY, Kang W, Kim EY. The significance of the trial outcome was associated with publication rate and time to publication. Journal of Clinical Epidemiology. 2017;84:78-84.
5. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Available from https://training.cochrane.org/handbook: Cochrane; 2020.
6. Kirkham JJ, Altman DG, Chan AW, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. BMJ. 2018;362:k3802.
7. Page MJ, Higgins JPT, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Available from https://training.cochrane.org/handbook: Cochrane; 2020.
8. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open. 2018;8:e019703.
9. Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying and adjusting for dissemination bias in meta-analysis are described. Journal of Clinical Epidemiology. 2016;80:25-33.
10. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. Systematic Reviews. 2017;6(1):204.
11. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. Systematic Reviews. 2016;5(1):108.

12. Page M, Sterne J, Higgins J, on behalf of the development group for ROB-ME. The ROB-ME (Risk Of Bias due to Missing Evidence) tool: a new tool for assessing reporting biases in evidence syntheses. Abstracts of the 25th Cochrane Colloquium, Edinburgh, UK. Cochrane Database of Systematic Reviews. 2018(9 Suppl 1):22-3. https://doi.org/10.1002/14651858.CD201801.
13. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. Journal of Clinical Epidemiology. 2017;86:39-50.
14. Chiocchia V, Nikolakopoulou A, Higgins JPT, Page MJ, Papakonstantinou T, Cipriani A, et al. ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. BMC Medicine. 2021;19(1):304.
15. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
16. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
17. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. Journal of Clinical Epidemiology. 2017;87:4-13.
18. Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Available from https://training.cochrane.org/handbook: Cochrane; 2020.
19. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Available from https://training.cochrane.org/handbook: Cochrane; 2020.
20. Hunter KE, Webster AC, Page MJ, Willson M, McDonald S, Berber S, et al. Searching clinical trials registers: guide for systematic reviewers. BMJ. 2022;377:e068791.
21. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. Journal of Clinical Epidemiology. 2008;61(10):991-6.

22. Marks-Anglin A, Chen Y. A historical review of publication bias. Research Synthesis Methods. 2020;11(6):725-42.
23. Vevea JL, Coburn K, Sutton A. Publication bias. In: Cooper H, Hedges LV, Valentine JC, editors. The Handbook of Research Synthesis and Meta-Analysis: Russell Sage Foundation; 2019. p. 383-430.
24. Maier M, VanderWeele TJ, Mathur MB. Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. Campbell Systematic Reviews. 2022;18(3):e1256.
25. Moreno SG, Sutton AJ, Turner EH, Abrams KR, Cooper NJ, Palmer TM, et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ. 2009;339:b2981.
26. Moreno SG, Sutton AJ, Thompson JR, Ades AE, Abrams KR, Cooper NJ. A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. Statistics in Medicine. 2012;31(14):1407-17.
27. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
28. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.
29. Mayo-Wilson E, Li T, Fusco N, Bertizzolo L, Canner JK, Cowley T, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. Journal of Clinical Epidemiology. 2017;91:95-110.
30. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Available from https://training.cochrane.org/handbook: Cochrane; 2020.
31. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890.
32. Lundh A, Boutron I, Stewart L, Hrobjartsson A. What to do with a clinical trial with conflicts of interest. BMJ Evidence-Based Medicine. 2020;25(5):157-8.
33. Gravier FE, Smondack P, Prieur G, Medrinal C, Combret Y, Muir JF, et al. Effects of exercise training in people with non-small cell lung cancer before lung resection: a systematic review and meta-analysis. Thorax. 2022;77(5):486-96.

Box 1. Possible scenarios in which study results are missing (adapted from Kirkham et al. (6))

Examples of scenarios where it is reasonable to suspect, given a lack of explanation from the study investigators, that a result is missing because of the P value, magnitude or direction of the result itself:

- study authors report in the Methods section, trials register entry, protocol or elsewhere that they measured (or intended to measure) the outcome of interest, but results are missing for the outcome;
- all results for an outcome were statistically non-significant and are reported incompletely (e.g. described only as "results were not significant", or means were reported with no measure of precision, or within-group change scores for the experimental group were reported but no data for the comparison group were presented), whereas results for other outcomes that were statistically significant are reported completely;
- results are missing for one of two outcomes that tend to be measured together (e.g. results are available for cause-specific mortality and are favourable to the experimental intervention, yet results for all-cause mortality, which must have been assessed given cause-specific mortality was, are missing);
- study authors pre-specified that they would report results separately for different outcomes (e.g. myocardial infarction, stroke, hypertension) yet instead report results for a composite outcome (e.g. "cardiovascular events") which happens to be statistically significant and favourable to the experimental intervention;
- summary statistics (number of events, or mean scores) are available only globally across all groups (e.g. study authors state that 10 of 100 participants in the study experienced adverse events, but do not report the number of events by intervention group); and
- a result is expected to have been generated for an outcome but it is not available and there is notable concern about conflicts of interest of primary study investigators or funders involved in the analysis or reporting, which have likely influenced them to withhold results that are unfavourable to the intervention (an assessment using TACIT ("Tool for Addressing Conflicts of Interest in Trials") (32) for the study should facilitate this judgement).

Examples of scenarios where it is reasonable to assume that a result is missing for a reason *unrelated* to the P value, magnitude or direction of the result include:

- it is clear that the outcome of interest was not measured in the study based on examination of the study protocol or statistical analysis plan or correspondence with the authors/sponsors;
- it can be assumed that the outcome of interest was not measured in the study because the instrument or equipment needed to measure the outcome was not available at the time or location where the study was conducted;
- the outcome of interest was measured but data were not analysed at all due to a reason unrelated to the nature of the results (e.g. there was a fault in the measurement instrument, funding for the research team discontinued, study staff changed jobs);
- study authors state that results for all outcomes measured appear in an appendix, but the appendix has been removed (or was not uploaded) by accident;
- study authors report results in a format unsuitable for inclusion in a meta-analysis for a reason unrelated to the P value, magnitude or direction of the result (e.g. study investigators report median and (interquartile) range for a continuous outcome because the data were skewed).

Table 1. Different reporting scenarios and the risk-of-bias tool – either a tool for assessing risk of bias due to missing evidence in a meta-analysis or a tool for assessing risk of bias in selection of the reported result in a study – which should be used to assess them (note that not all represent a high risk of bias).

Scenario	Tool to assess risk of bias due to missing evidence (e.g. ROB-ME)	Tool to assess risk of bias in selection of the reported result (e.g. RoB 2)
Study authors pre-specify in ClinicalTrials.gov that pain will be measured yet present no result for pain in any report.	✓	
Study is listed on a pharmaceutical company's clinical study register, and several outcomes of interest (such as "depression measured using the Beck Depression Inventory") are pre-specified in the abstract for the protocol. However, no results are publicly available, and the request for the required results was rejected by the company.	✓	
Study authors state in the Methods section that "health-related quality of life" was measured, but only state in the Results section that there was "no significant difference between groups in quality of life". Study authors do not respond to requests for fully reported results for this outcome.	✓	
A journal article published in 2015 describing results of a trial for rheumatoid arthritis has no results for "swollen joints". However, measurement of swollen joints is expected to have occurred because it has been deemed a patient-important outcome recommended for assessment in all trials since 1994, when it was included in a core outcome set for rheumatoid arthritis.	✓	
Study authors report in the Methods section that they measured depression using three instruments (HAM-D, BDI and MADRS) at two time points. However, results are reported for one of the instruments (BDI) at one time point only, and authors provide no further information about results for the other measures of depression. The systematic reviewers are only willing to include results for the HAM-D measure of depression in the meta-analysis, thus there is no result available for inclusion.	✓	
Study authors report in the Methods section that they measured anxiety using three instruments at two time points. However, results are reported for one of the instruments at one time point only, and authors provide no further information about results for the other measures of anxiety. The systematic reviewers are willing to include results for any measure of anxiety in the meta-analysis, thus there is a result available for inclusion.		✓
Study authors report that they conducted multiple analyses, each adjusted for different prognostic factors. However, only the unadjusted effect estimate was fully reported, and all adjusted results are referred to only as being "not significant (data not shown)".		✓
Study authors pre-specify a cut-point on a continuous measurement scale to create categories of "improved" versus "not improved", yet the reported result is based on a different cut-point that was selected post-hoc.		✓
Study authors report results for "active range of motion in flexion" in the journal article, yet this outcome was not pre-specified in the publicly available trial protocol.		✓

Table 2. Summary of the issues addressed in the ROB-ME tool

Section	Issues addressed*
Within-study assessment of non-reporting bias ('known unknowns')	Whether, of the studies identified: • there was any for which no result was available for inclusion in the meta-analysis, likely because of the P value, magnitude or direction of the result generated • (if applicable) it is likely that there would be a notable change to the summary effect estimate if the omitted results had been included • there was any for which it was unclear whether an eligible result was generated • (if applicable) it is likely that there would be a notable change to the summary effect estimate if the potentially omitted results had been included
Across-study assessment of non-reporting bias ('unknown unknowns')	Whether: • circumstances indicate potential for some eligible studies not being identified because of the P value, magnitude or direction of the results generated • (if applicable) it is likely that studies not identified had results that were eligible for inclusion in the meta-analysis • (if applicable) the pattern of observed study results suggest that the meta-analysis is likely to be missing results that were systematically different (in terms of P value, magnitude or direction) from those observed • (if applicable) sensitivity analyses suggest that the summary effect estimate was biased due to missing results

*For the precise wording of signalling questions and guidance for answering each one, see the full risk-of-bias tool at https://www.riskofbias.info/.

Figure 1. Summary of the process of assessing risk of bias due to missing evidence in meta-analyses

Figure 2. Example forest plot displaying results missing from a meta-analysis of the effect of preoperative exercise training compared with usual care on postoperative complications (data from Gravier et al. (33))