Using Risk of Bias 2 to assess results from randomised controlled trials: guidance from Cochrane

Ella Flemyng,1 Theresa Helen Moore,2,3 Isabelle Boutron,4,5 Julian PT Higgins,2,3 Asbjørn Hróbjartsson,6,7 Camilla Hansen Nejstgaard,6,7 Kerry Dwan 8,9

Abstract
A systematic review identifies, appraises and synthesises all the empirical evidence from studies that meet prespecified eligibility criteria to answer a specific research question. As part of the appraisal, researchers use explicit methods to assess risk of bias in the results’ from included studies that contribute to the review’s findings, to improve our confidence in the review’s conclusions. Randomised controlled trials included in Cochrane Reviews have used a specific risk of bias tool to assess these included studies since 2008. In 2019, a new version of this tool, Risk of Bias 2 (RoB 2), was launched to improve its usability and to reflect current understanding of how the causes of bias can influence study results. Cochrane implemented RoB 2 in a phased approach, with users of the tool informing guidance development. This paper highlights learning for all systematic reviewers (Cochrane and non-Cochrane) from the phased implementation, highlighting differences between the original version of the tool and RoB 2, consideration of reporting systematic review protocols or full review reports that have used RoB 2, and some tips shared by authors during the pilot phase of the implementation.

Introduction
A systematic review identifies, appraises and synthesises all the empirical evidence from studies that meet prespecified eligibility criteria to answer a specific research question. Researchers conducting systematic reviews seek to use explicit, systematic methods that are selected with a view aimed at minimising bias and imprecision to inform decision making. Assessing risk of bias in the design, conduct and reporting of the included studies in the review is a key step. This helps systematic reviewers, and the users of these reviews, to understand whether there is any risk of bias in the included study’s results that could distort the review’s results. This helps improve our confidence in the review’s conclusions and that it accurately represents the truth (that the true effect of the intervention has not been overestimated or underestimated).

Cochrane defines bias in this context as ‘a systematic error, or deviation from the truth, in results’. Before the mid-2000s, tools that critically assessed studies within reviews typically considered the broader notion of ‘methodological quality’ which often involved a combination of risk of bias, imprecision, relevance, applicability, ethics and completeness of reporting. However, this breadth of concepts was not based on any particular underlying theoretical framework for ‘quality’ assessment and therefore different tool resulted in different assessments for the same study. To address this, in 2008, Cochrane released a risk of bias tool to focus on the single concept of risk of bias in randomised controlled trials (RCTs) included within its reviews of interventions. Other concepts are now considered at...
other stages in the review process. For example, precision (the extent to which results are free of random errors) and external validity (directness, applicability or generalisability) are domains of GRADE (Grading of Recommendations, Assessment, Development and Evaluations).3

In 2019, a new version of the tool, Risk of Bias 2 (RoB 2), was launched to improve its usability, address some of the limitations of the original version3 5 and to reflect current understanding of how the causes of bias can influence study results.5 7 8 A recent study by Minozzi et al9 found that 69% (95/137) reviews published between 2019 and 2021 adhered to RoB2 by assessing an outcome rather than the study. However, the majority of these reviews only included a single primary outcome and when the assessment was restricted to reviews with at least two primary outcomes, adherence decreased to 29%.

Cochrane used a phased implementation approach to introduce RoB 2 to authors of its systematic reviews. It aimed for a supported and gradual roll-out to observe and address common issues as they arose.8 Following the publication of the first Cochrane Review from the pilot phase,10 RoB 2 is now the recommended tool for assessing risk of bias in RCTs in Cochrane Reviews, though the original risk of bias tool is still acceptable (https://community.cochrane.org/news/status-and-expectations-implementation-risk-bias-2-cochrane-intervention-reviews).

Over 150 Cochrane Reviews are now in development that use RoB 2. In this paper, the team that led the implementation and members of the Cochrane Bias Methods Group share the guidance developed to support authors using RoB 2 in systematic reviews.

Online supplemental appendix 1 details the key resources, training, tools, software and template recommended by Cochrane that are available for systematic review authors to help them learn about and use RoB 2.

Differences between the original risk of bias tool and RoB 2

A comparison of the differences between the two tools can be seen in table 1. One key difference between the two is that risk of bias assessments for RoB 2 are more explicitly associated with a specific result for an outcome from each included study (result-level assessment).

This recognises that different outcomes from one study could have different issues that affect the risk of bias, for example, an objective outcome (eg, maximal cardiorespiratory fitness) versus a subjective outcome (eg, health-related quality of life) as seen in Williams et al14 Cochrane Review on physical activity interventions for people with congenital heart disease. For one of the studies (‘Sandberg 2018’), the authors judged the health-related quality of life result to be of high risk of bias because the participants both reported the outcome and were aware of the intervention, and they felt that this knowledge would have likely influenced the outcome (RoB 2, domain 4; blinding of the outcome assessment). In contrast, they judged this same domain in the same study to be at low risk of bias for the maximal cardiorespiratory fitness result because most of the assessors were blinded and for those who were not blinded, the authors felt that this knowledge would not have influenced the measurement as it is an objective outcome.

In addition, RoB 2 recognises that different results for the same outcome from one study could have different issues that affect the risk of bias. As an example, a Cochrane Review by Kew et al15 on increasing the dose of inhaled corticosteroids when asthma symptoms worsen to reduce the need for further treatment included the outcome of treatment failure. This outcome was assessed in two populations; all randomised participants and those participants who started their study inhaler. The result for each outcome (treatment failure in all randomised participants vs treatment failure in only those who took their study inhaler) were different and

<table>
<thead>
<tr>
<th>Table 1 Differences between the original risk of bias tool and RoB2 for assessing randomised controlled trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original risk of bias tool</strong></td>
</tr>
<tr>
<td><strong>Focus of assessment</strong></td>
</tr>
<tr>
<td><strong>Structure</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Domains</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Basis of judgement</strong></td>
</tr>
<tr>
<td><strong>Judgement options</strong></td>
</tr>
</tbody>
</table>

Guidance for using the original risk of bias tool is available in V.5.2 of the Cochrane Handbook.

*As RoB 2 assesses risk of bias in a numerical result, risk of bias due to missing results (from non-publication of trials or results) is not part of the revised tool. This can be assessed using the Risk of Bias-Missing Evidence; available on the Risk of Bias tools website (https://www.riskofbias.info/welcome/rob-me-tool).

†RoB 2 does not include an ‘other bias’ domain. All domains within RoB 2 include trial features that directly influence the risk of bias; those that indirectly influence the risk of bias should be assessed in other stages of the review, or example, for conflicts of interest A Tool for Addressing Conflicts of Interest in Trials is in development.16 RoB2, Risk of Bias 2.
for each result there were different issues that affected the risk of bias, as seen in figure 1. For one of the studies (‘Oborne 2009’), the authors judged the treatment failure in all randomised participants result to be of low risk of bias because all participants contributed data for the intention-to-treat analysis up to the point at which they left the study (RoB 2 domain 3; missing outcome data). In contrast, they judged this same domain in the same study to be of high risk of bias for the treatment failure in only those who took their study inhaler result as this population represents a relatively small, non-randomised proportion of the full cohort, there was an imbalance in the number of people in each group who took the study inhaler, and patients who discontinued who did not use the study inhaler might have done so for reasons relating to their disease worsening. Another hypothetical example is a systematic review interested in a outcome measured at end of treatment, such as 4 weeks, and at longest follow-up, such as months or years later. These two results for the same outcome may theoretically have different risk of biases, that is, a lower risk of bias at end of treatment and higher risk of bias at longest follow-up due to missing outcome data.

Another important development is that RoB 2 requires authors to specify whether the result being assessed is being interpreted as the effect of assigning participants to interventions (the intention-to-treat effect) or the effect of participants adhering to their assigned intervention according to the trial protocol (the per-protocol effect). Authors need to think carefully about the aim of their systematic review, and which is most appropriate for each result. The effect of interest will have implications for the risk of bias due to deviations from the intended intervention (RoB 2, domain 2). An example of this is again seen in Kew et al Cochrane Review where, in the included trials, the randomised population only starts the intervention at the onset of treatment failure. The critical outcome for the review was treatment failure in all randomised participants (the intention-to-treat effect), and an important secondary outcome was treatment failure in participants using the study inhaler (the per-protocol effect); the same outcome in different populations resulted in different effects of interest and different risk of bias assessments (see figure 1).

By considering the specific result for each outcome of interest from a study separately, we can more accurately assess bias for those results which will increase our confidence in the evidence presented in the systematic review.

Richter and Hemmingsen compared RoB 2 to the original tool and found that the mean assessment times were comparable and there were few difficulties achieving agreement between raters with both tools. The biggest divergence between the tools was with subjective outcomes in open-label studies, where the original tool was more likely to penalise than RoB 2. The original tool led to harsher judgements due to the options available (high/low/unclear risk) whereas the signalling questions and guidance made RoB 2 easier to work through complexity and context.

This paper concentrates on guidance for parallel RCTs only. The RoB 2 tool has two supplemental variants, one for cluster RCTs and one for crossover RCTs, both available from the Risk of Bias tools website (https://www.riskofbias.info/welcome/rob-2-0-tool).

RoB 2 considerations for systematic review protocol development

It is important to consider carefully at the protocol stage how RoB 2 will be used in a review. As RoB 2 is a result-level tool and does not need to be completed for all the outcomes and all the studies in the systematic review, prioritisation of which outcomes to assess is likely to be needed. Without considering this, reviewers can increase their workload substantially and are at risk of using the tool incorrectly, that is, with a study-level focus instead of the correct results-level focus. Cochrane recommends that review authors limit the RoB 2 assessments to the results for their outcomes planned for inclusion in a summary of findings table, that is, those outcomes deemed to be the most critical and important outcomes for decision making. Table 2 details these considerations.

Rows highlighted green are core to the context of a systematic review and must be prespecified to facilitate the accurate use of
RoB 2 during risk of bias assessments. Includes guidance from chapter 7 and chapter 8 of the Cochrane Handbook, as well as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020.

RoB 2 considerations for reporting the full systematic review

Table 3 details the considerations needed when reporting the full systematic review.

Note that this checklist only highlights RoB 2 considerations for review reporting. Rows highlighted green are core to understanding the risk of bias assessments in the context of a systematic review's conclusions. Some journals may have specific guidance on use of tables and figures, so this guidance may need to be adapted to meet those journal requirements. Includes guidance from chapter 7 and chapter 14 of the Cochrane Handbook.
RoB 2 tips
In this section, we bring together some of the key takeaways from the RoB 2 pilot phase of implementation.

Create a RoB 2 decision tool following protocol development
While RoB 2 is a result-based assessment, considering which domains are expected to be consistent across results within a study and designing the data collection form accordingly can save a lot of time. This early investment goes a long way. Some teams have created a risk of bias decision tool that is specific to their review, to help reviewers make consistent decisions and to ease the process of assessing bias, for example, issues in randomisation will be common to all outcomes, issues of missing data may differ for outcomes at different time points, and issues of outcome assessment may be different between patient-reported outcomes and outcomes derived from routine data sources. The first few assessments may take some time to get right but once done, subsequent assessments naturally become much easier and faster.

Disagreements are no bad thing
Practising a couple of assessments will always highlight differences that can be ironed out, but inter-rater discrepancies beyond that should be expected and may even improve the review. The signalling questions in RoB 2 provide a clearer framework for discussing differences in judgements and justifications than the old tool, and the process of doing so is a key part of gaining understanding and interrogating the evidence.

---

Table 3  Seven items Cochrane recommends authors report in a completed systematic review that used the RoB 2 tool

<table>
<thead>
<tr>
<th>What to report</th>
<th>Further details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods—‘Assessment of risk of bias in included studies’</td>
<td></td>
</tr>
<tr>
<td>1. Include all the RoB 2 considerations from the Protocol.</td>
<td>Ensure the review includes all the RoB 2 information from the protocol. If there were any deviations from the protocol, these should be detailed in the review’s methods section, along with justification.</td>
</tr>
<tr>
<td>2. State the version of the RoB 2 tool that was used.</td>
<td>The riskofbias.info website lists the current version and archived versions of the RoB 2 tool. Ensure that the version used is stated.</td>
</tr>
<tr>
<td>Results—‘Risk of bias in included studies’</td>
<td></td>
</tr>
<tr>
<td>3. Refer to results level RoB 2 tables, which includes the support for judgement for each domain assessment.</td>
<td>Tables summarising the risk of bias for all results that contribute to an outcome should be included in the review. Each result for the outcomes prespecified for risk of bias assessments (likely to be the reviews’ critical and important outcomes) should have a table that includes the risk of bias judgements (high, low or some concerns) and the support each judgement. In certain circumstances, authors may wish to use other figures that best present the risk of bias data, for example, weighted risk of bias bar plots can provide a succinct summary when there are lots of studies in a synthesis. See figure 6 for an example of a risk of bias figure for one result for one outcome created using robvis.</td>
</tr>
<tr>
<td>4. State how to access detailed risk of bias assessments data (with consensus responses to the signalling questions).</td>
<td>Authors should state how the data can be accessed, for example, via an online supplemental file, repository or other.</td>
</tr>
<tr>
<td>5. Provide a brief overview of the risk of bias assessments.</td>
<td>Consider overall comments on key aspects of the risk of bias assessments, for example, the quality of randomisation and extent to which blinding was implemented. Consider whether there are important differences in risk of bias by outcome. If risk of bias assessments are very similar (or identical) for all outcomes in the review, a summary of the assessments across studies should be presented here. If risk of bias assessments are very different for different outcomes, this section should be very brief, and summaries of the assessments across results should be discussed with other GRADE considerations in the Discussion (see point 7 below).</td>
</tr>
<tr>
<td>Results—Describing the effects of interventions</td>
<td></td>
</tr>
<tr>
<td>6. Refer to visual representations of the risk of bias assessments in relation to each result.</td>
<td>Using forest plots with traffic lights is highly recommended (see Figure 2-6 in Williams et al10). For synthesis without meta-analysis, we recommend that a column is added to any visual representation of the data to highlight the overall risk of bias associated with each of the results in the table/figure (see Table 1 in Williams et al10). Guidance: section 7.6 of the Cochrane Handbook.</td>
</tr>
<tr>
<td>Results—‘Subgroup analysis’</td>
<td></td>
</tr>
<tr>
<td>(If applicable) Discuss any subgroup analysis conducted that relates to the overall risk of bias judgements.</td>
<td>When preparing subgroup analyses it is advisable to use the ‘overall’ judgement of bias, rather than using bias relating to specific domains. Guidance: section 7.6.2 of the Cochrane Handbook.</td>
</tr>
<tr>
<td>Results—‘Sensitivity analysis’</td>
<td></td>
</tr>
<tr>
<td>(If applicable) Discuss any sensitivity analysis conducted that relates to the overall risk of bias judgements.</td>
<td>When preparing sensitivity analyses it is advisable to use the ‘overall’ judgement of bias, rather than using bias relating to specific domains. Guidance: section 7.6.2 of the Cochrane Handbook.</td>
</tr>
<tr>
<td>Discussion—‘Certainty of the evidence’ (previously the ‘Quality of the evidence’ section)</td>
<td>Along with the other GRADE considerations, highlight any important implications from the risk of bias assessments for each of the outcomes prespecified for risk of bias assessments. These are likely to be the reviews’ critical and important outcomes or those included in the summary of findings table. For example, if the risk of bias assessments results in downgrading the certainty of the evidence for a specific outcome and whether the effects of the intervention need to be interpreted with caution. Guidance: section 7.5 and 14.2.2 of the Cochrane Handbook.</td>
</tr>
</tbody>
</table>

GRADE, Grading of Recommendations, Assessment, Development and Evaluations; RoB 2, Risk of Bias 2.
Back to bias assessment as it was always intended

Shifting from assessing studies to assessing results may initially feel like a daunting task but, once a rhythm is found, it can refocus the mind on why bias assessment is so important in systematic reviews. RoB 2 provides a framework for building meaningful bias considerations through reviews, from protocol planning to writing up results and implications for practice.

Authors are not expected to assess risk of bias for all results from all included studies

The risk of bias assessment should focus on results of studies that contribute information to outcomes that users of the review will find most useful. This will generally correspond to the results that are used to populate outcomes in summary of findings tables; however, this will depend on the review question and protocol, which may have specified other outcomes for risk of bias assessment.

Conclusions

RoB 2 is the recommended tool to assess the risk of bias of specific results included in a systematic review. It has some fundamental differences from the original Cochrane risk of bias tool and these need to be understood before the tool is applied. This includes whether the authors are interested in the effect of assignment or adherence and which results they plan to assess. Prespecifying some items from RoB 2 can aid implementation of the tool. Transparency is key and all supporting material should be submitted alongside the full review and for peer review. An Excel tool is available on the risk of bias website (www.riskofbias.info) to facilitate completion of the assessments. Assessing the risk of bias for specific results allows us to see the impact this has on a synthesis or meta-analysis, whereas previously a study was usually judged for bias overall and might have left some results with a worse assessment of bias than was warranted.

Author affiliations

1Cochrane Evidence Production and Methods Directorate, Cochrane, London, UK
2Population Health Sciences, University of Bristol Medical School, Bristol, UK
3NIHR Applied Research Collaboration West at University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
4Inserm, INRAE, Centre de Recherche Epidemiologie et StatistiqueS Université Paris Cité, Université Paris Cité and Université Sorbonne Paris Nord, Paris, France
5Centre d’Epidémiologie Clinique, AP-HP, Hôpital Hôtel Dieu, Paris, France
6Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark
7Open Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark
8Evidence Production and Methods Directorate, Cochrane, London, UK
9Centre for Reviews and Dissemination, University of York, York, UK

Twitter Kerry Dwan @dwan_kerry
Contributors KD conducted the study by leading the RoB2 pilot and wrote the guidance, commented on the manuscript, submitted the manuscript and acts as guarantor for the manuscript. EF conducted the study by leading the RoB2 pilot and wrote the guidance and this manuscript. THM conducted the study, wrote the guidance and commented on the manuscript. JPTH was involved in the conduct of the study, contributed towards the guidance and commented on the manuscript. IB was involved in the conduct of the study, contributed towards the guidance and commented on the manuscript. AH was involved in the conduct of the study, contributed towards the guidance and commented on the manuscript. CHN commented on the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests EF is employed by Cochrane. During the development of the manuscript, KD and THM were employed by Cochrane. JPTH, IB, AH and CHN are leads of the Cochrane Bias Methods Group. JPTH, IB and AH were part of the development team for the RoB 2 tool.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Not applicable.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Theresa Helen Moore http://orcid.org/0000-0003-3719-6361
Kerry Dwan http://orcid.org/0000-0001-6918-1215

References