Development of ROBUST-RCT: Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials

Ying Wang,[1,2] Sheri Keitz,[3] Matthias Briel,[2,4] Paul Glasziou,[5] Romina Brignardello-Petersen,[2] Reed A C Siemieniuk,[2] Dena Zeraatkar,[2,6] Elie A Akl,[2,7] Susan Armijo-Olivo,[8,9] Dirk Bassler,[10] Carrol Gamble,[11] Lise Lotte Gluud,[12] Jane Luise Hutton,[13] Luz M Letelier,[14] Philippe Ravaud,[15] Kenneth F Schulz,[16] David J Torgerson,[17] Gordon H Guyatt[2,18,19]

For numbered affiliations see end of the article

Correspondence to: Y Wang
yingwwy@163.com
(or @yingwwy on X;
ORCID 0000-0002-2296-1991)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2025;388:e081199
http://dx.doi.org/10.1136/bmj-2024-081199

Accepted: 27 February 2025

Recent innovations in evidence based medicine methods, in particular instruments assessing risk of bias in randomised trials, have focused on methodological rigour at the expense of simplicity and practicability. Such a focus could lead to challenges in application and loss of reliability of instruments. To deal with these shortcomings, the Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials (ROBUST-RCT) was created—a rigorously developed, simply structured, and user friendly instrument for assessing risk of bias of randomised controlled trials included in systematic reviews. This paper describes the development of ROBUST-RCT and provides associated documents and a manual of instructions.

SUMMARY POINTS

ROBUST-RCT (Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials) is a rigorously developed, simply structured, and user friendly instrument for assessing risk of bias of randomised controlled trials in systematic reviews

The aim of ROBUST-RCT is to achieve an optimal balance between simplicity and methodological rigour

Systematic review teams with different levels of expertise can use ROBUST-RCT when undertaking risk of bias assessments

Although systematic reviews of randomised controlled trials provide the best evidence for the effects of healthcare interventions,[1] flaws in trial design and conduct may result in biased estimates of effects, and hence misleading conclusions.[2] As a result, risk of bias assessment of randomised controlled trials has become an essential step in the systematic review process. Furthermore, risk of bias represents one domain in the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system of rating certainty of evidence, and trial limitations resulting in risk of bias may lead authors of systematic reviews to rate down the certainty of evidence.[3 4]

Although many instruments for assessing risk of bias in randomised controlled trials are available,[5] most have important limitations. A systematic survey found that existing instruments often include items that do not deal with risk of bias.[5] To be suitable for use in systematic reviews, risk of bias instruments should include only items that deal with risk of bias problems rather than other GRADE domains.[3]

The most popular and rigorously developed instruments include those offered by the Cochrane Collaboration. The first Cochrane risk of bias instrument[6] included an "unclear" response option that failed to take advantage of reasonable inferences about the presence or absence of risk of bias.[7] Users of this instrument have reported problems with assessing the incomplete outcome data and the selective reporting domains.[8]

The revised Cochrane instrument for assessing risk of bias in randomised controlled trials, RoB 2,[9] intended to replace the first instrument, introduced non-intuitively labelled domains and a less than straightforward series of signalling questions and algorithms for assessing each domain. The sophisticated algorithms (up to seven signalling questions)[10] and difficulty in understanding new terminologies—for example, "deviations from the intended intervention that arose because of the trial context"—raised challenges for systematic reviewers.[11]

Possibly as a consequence of these limitations, uptake of RoB 2 is relatively low in non-Cochrane reviews, and misapplication is common.[12 13] Previous published studies have documented the low interrater reliability of RoB 2 and documented its challenges in implementation, even when used by systematic reviewers with substantial expertise.[14 15] In particular, less experienced systematic reviewers—those in systematic review teams who often assess risk of bias of individual randomised controlled trials—may experience daunting challenges in applying RoB 2.[11]

In considering the possibility of developing a new instrument that deals with the limitations of RoB 2, we contacted nine international experts who were well published in the area of risk of bias assessment in randomised controlled trials. These individuals agreed on the limitations of RoB 2 related to its complexity, and they shared the experience of the challenges that the less experienced members of their systematic review team faced in applying the instrument.

We have argued that movements in clinical epidemiology and evidence based medicine have lost sight of the optimal balance between simplicity and methodological rigour, with RoB 2 representing one example.[11] This perspective motivated us to use rigorous methodology, while bearing simplicity in mind, to develop a new instrument. We aimed to create an instrument to serve the needs of systematic review teams with less experienced members assessing risk of bias. This paper describes the development of the Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials (ROBUST-RCT).

Methods

The instrument development team

Operations committee

Members of the operations committee (GG, YW, RBP, RAS, DZ) identified the need for a new instrument, developed a protocol (see supplementary appendix 1), recruited the panel of experts, organised materials, presented proposals to the panel, and created drafts of the instrument and associated materials.

Panel

The operations committee identified experts in risk of bias assessment from the author lists of methodological papers that stated or indicated that they dealt with risk of bias. By screening the references of existing risk of bias instruments of randomised controlled trials and their guidance documents (by conducting a systematic survey),[5] as well as eligible papers suggested by members of the operations committee, we identified the first round of eligible papers. Then we screened the references of these papers to identify additional eligible papers. We identified 295 eligible papers in total. Individuals eligible for panel membership had participated as first, last, or corresponding authors of at least one eligible paper, and as coauthor of at least two other papers. From a total of 63 eligible experts, stratified by region and sex, we randomly selected 10 and invited them to join the panel; nine agreed. The panel included two more methodological experts (MB, PG) who the committee members knew and thought could make substantial contributions.

In addition, members of the operations committee suggested a list of 22 internationally recognised expert educators in evidence based medicine, from whom we randomly selected two, stratified by sex, to join the panel; both agreed. The panel included a third individual (SK) known to committee members as an exceptionally astute educator in evidence based medicine. The three educators came from different regions.

The panel included 19 members: five individuals from the operations committee and 14 additional members. Sixteen members had expertise in methodology of risk of bias assessment (GG, YW, RBP, RAS, DZ, MB, PG, EAA, SAO, DB, CG, LLG, JLH, PR, KFS, DJT) and three were experienced educators in evidence based medicine (SK, RJ, LML). This international collaboration included 10 men and nine women: seven from North America, five from Europe, three from the UK, two from Oceania, one from South America, and one from Asia.

Ground rules for instrument development

Seven ground rules developed by the operations committee and endorsed by the panellists, guided the instrument development process:

- The instrument aims to assess risk of bias of randomised controlled trials in the context of systematic reviews.
- The objective is to develop a user friendly instrument: item presentation will be simple and straightforward; making judgments not overly complex or difficult.
- We define bias as a systematic error or systematic deviation from the truth.
- We assume that systematic reviewers will use the GRADE approach to assess certainty of evidence.
- Decisions should be consistent with the GRADE system in distinguishing risk of bias from imprecision (random error), indirectness (applicability), and publication bias. Reporting quality represents another concern to distinguish from risk of bias.
- The instrument currently deals only with risk of bias assessment of individually randomised parallel group trials. The risk of bias assessment of cluster trials and crossover trials is for future consideration.
- This instrument will not include items for the detection of fraud.

Collection of candidate items

To collect candidate items, we systematically surveyed the 17 risk of bias instruments of randomised controlled trials published from 2010 to October 2021 for the included items (see details in a separate publication).[5] We extracted additional candidate items from two studies: one study collected items that Cochrane reviewers regarded as "other bias" when they used the Cochrane's first risk of bias instrument,[16] and the other study summarised the published comments on the Cochrane Collaboration's first instrument.[17]

Through a survey of item classification in which 13 panellists participated and judged what concern the items addressed (risk of bias, imprecision, indirectness, reporting quality, or none of the aforementioned), we classified the items into three categories[5]: category 1 included items that the majority of the panellists judged as addressing risk of bias, category 2 included items that the majority of the panellists judged as not addressing risk of bias, and category 3 included items that generated substantial disagreement among the panellists about whether the items addressed risk of bias.

To generate an organised item list for efficient discussion by the panel, the operations committee combined the highly related items (for example, items dealing with different aspects of missing outcome data). We removed items that specifically dealt with problems relevant to cluster or crossover trials.

Empirical evidence from meta-epidemiological studies

To provide empirical evidence for item selection for our instrument, we conducted a systematic survey of meta-epidemiological studies examining the impact of potential risk of bias problems (items in categories 1 and 3) on effect estimates in randomised controlled trials.[18] A separate paper presents the methods and results.[18]

Panel process

The operations committee presented issues to discuss and proposals to the panel. Panel meetings, co-chaired by YW and GG, used an open discussion format in which panellists first spoke freely, after which GG guided the panel towards consensus. After each meeting, YW produced minutes including the panel's tentative decisions and the discussion involved. Panellists revisited controversial topics in subsequent meetings. Through 16 1.5-hour panel meetings and associated email conversations from February to October 2023, the panel achieved consensus on item selection, instructions for included items, and format of the instrument.

The operations committee presented the organised item list to the panel. The panel discussed each item in category 2 (in which items were judged as not addressing risk of bias by majority of panellists), then category 1 (in which items were judged as addressing risk of bias by majority of panellists), and finally category 3 (substantial disagreement among panellists about whether the items addressed risk of bias or not).

The panel used six criteria for item selection (box 1) developed by the operations committee and endorsed by the panel to help decisions on items in categories 1 and 3. No single criterion or group of criteria were deemed essential. The more criteria an item met, the more likely it was to be suitable for selection as an item in the instrument.

The panel chose core items for the instrument. The panel also identified items of potential importance that although rejected as core items were ultimately chosen as optional items for the instrument.

The operations committee drafted instructions for core items and considerations about optional items. The panel discussed and revised the draft and approved the final version. We developed a manual to support the instrument's use.

Box 1: Six criteria for item selection

- Clearly a risk of bias problem rather than imprecision, indirectness, publication bias, or reporting quality
- Theoretical or logical argument for why the item is important
- Information required to make judgment on the item is commonly reported in trials
- Non-expert systematic reviewers can make the judgment easily
- Problem occurs more often than rarely
- Empirical evidence supports item influence on effect estimates

User testing exercises

To identify challenges experienced by junior systematic reviewers in comprehending and applying the instrument, we conducted user tests. We enrolled 15 people who had assessed risk of bias in randomised controlled trials for at least one systematic review and had never led a systematic review of randomised controlled trials. The participants varied in respect of sex, country, clinical background, student status, and number of systematic reviews of randomised controlled trials in which they had assessed risk of bias (see supplementary appendix 2). We identified eligible individuals through suggestions from panel members. Recruitment was discontinued once we had achieved saturation of comments on the instrument.

For user testing, the panellists suggested randomised controlled trials that presented challenges in risk of bias assessment. Two committee members (YW and GG) assessed risk of bias in these trials and then selected five trials (see supplementary appendix 2) in which systematic reviewers would face challenges in assessing as many items as possible in the instrument. We ensured that the trials presented challenges in each item.

Each participant received one trial, the draft of the instrument, and the manual. YW conducted a think-aloud interview of about one hour with each participant. During the interviews, participants applied the instrument to the trial and articulated the thought process for each item that led to their assessment. YW compared the participant's assessment with the assessment made and agreed on by YW and GG; when mistakes or problems occurred, YW explored the reasons. Participants expressed their overall experience in applying the instrument.

To identify concerns or questions that the systematic review experts might have about the instrument, we conducted a second user testing exercise. We searched the Cochrane Library, randomly selected Cochrane systematic reviews published between 1 January 2019 and 14 February 2024, and identified the first, last, or corresponding authors. If the authors had been the lead for at least five systematic reviews of randomised controlled trials (not limited to Cochrane reviews), we invited them to participate in user testing. The eight participants varied in respect of sex, country, and clinical background (see supplementary appendix 2).

Before the interviews, the review experts received the instrument and manual. To explore any concerns and suggestions, YW followed a semistructured interview guide, interviewing each participant for about one hour.

YW recorded and transcribed the interviews from both the user testing groups and extracted people's feedback, comments, and suggestions. GG and YW reviewed the results after completing interviews for each five junior systematic reviewers and after completing interviews for four, six, and eight review experts. Together they identified concerns and solutions and presented these to the panel in email communications, ultimately deciding on modifications

Table 1 | Initially selected core items and optional items and judgment about whether they met the six criteria for item selection*

Items	Item selection criteria					
	Clearly a risk of bias rather than other concerns†	Theoretical or logical argument for why item is important	Information required to make judgment is commonly reported in trials	Non-expert reviewers can make judgment easily	Problem occurs more often than rarely	Empirical evidence supports item influence on effect estimates‡
Initially selected core items						
Random sequence generation	Yes (category 1)	Yes	Yes	Yes	Yes	Overestimation (moderate certainty)
Allocation concealment	Yes (category 1)	Yes	Yes	Yes	Yes	Overestimation (moderate certainty)
Blinding of participants	Yes (category 1)	Yes	Yes	Yes	Yes	Any outcomes: very uncertain Patient reported outcomes: overestimation (moderate certainty) Observer reported or objective outcomes: very uncertain
Blinding of healthcare providers	Yes (category 1)	Yes	Yes	Yes	Yes	Very uncertain
Blinding of outcome assessors	Yes (category 1)	Yes	Yes	Yes	Yes	Any outcomes: very uncertain Objective outcomes: very uncertain Subjective outcomes: overestimation (high certainty)
Missing outcome data	Yes (category 1)	Yes	Yes	Yes	Yes	Underestimation (low certainty)
Intention-to-treat analysis§	No (category 3)	Yes	No	Yes	Yes	Very uncertain
Initially selected optional items						
Whether baseline prognostic factors were balanced between groups	No (category 3)	Yes	Yes	No	Uncertain	Very uncertain
Whether co-interventions were balanced between groups in blinded trials	No (category 3)	Yes	No	No	Uncertain	Overestimation (low certainty)
Whether outcome assessment or data collection differed between groups	Yes (category 1)	Yes	No	No	No	No evidence
Whether follow-up time, frequency, or intensity of outcome assessment differed between groups	Yes (category 1)	Yes	No	No	No	No evidence
Whether outcome measurement method was valid (ie, validity of outcome measurement)	No (category 3)	Yes	No	No	No	No evidence
Whether there was selective reporting	No (category 3)	Yes	No	No	No	Very uncertain
Whether the trial was terminated early for benefit	No (category 3)	Yes	Yes	Yes	Yes	Overestimation (moderate certainty)

*Supplementary appendix 3 summarises judgment about whether all items in categories 1 and 3 met the item selection criteria.
†Other concerns to distinguish with risk of bias include imprecision, indirectness, publication bias, and reporting quality.
‡Empirical evidence from a systematic survey of meta-epidemiological studies.[18]
§After user testing, the panel split the initial intention-to-treat analysis item into two concerns: per protocol analysis and as treated analysis. The panel combined the per protocol analysis concern with the missing outcome data concern as the ultimate core item 6, and added the as treated analysis concern as the optional item 6.

to the instrument and manual. Supplementary appendix 2 summarises the feedback and resulting changes. After each revision, subsequent user testing presented participants with the updated version of the instrument and manual.

Results

Panel's initial decisions

The panel selected items from a list containing 29 items: 10 items in category 1 (in a survey the majority judged the items as addressing risk of bias), nine in category 2 (majority judged the items as not addressing risk of bias), and 10 in category 3 (substantial disagreement about whether the items addressed risk of bias or not) (see supplementary appendix 3).

The panel initially selected seven core items (six from category 1 and one from category 3) and seven optional items (two from category 1 and five from category 3). Table 1 presents the extent to which these items met the six item selection criteria (see box 1). Supplementary appendix 3 summarises the panel's decisions for all items and rationale.

The panel initially developed two versions of the instrument (see supplementary appendix 3, tables 4 and 5). Version A asked the systematic reviewers assessing the individual trials to evaluate what happened in each trial for each item (eg, item 3, judge if participants were blinded). Response options included definitely yes, probably yes, probably no, and definitely no. Version B asked the systematic reviewers assessing the individual trials to decide the extent to which any deficits in instituting methodological safeguards resulted in risk of bias (eg, item 3, judge if failure to blind participants resulted in risk of bias). Response options for risk of bias included definitely low, probably low, probably high, and definitely high.

Revision based on user testing

User testing with junior systematic reviewers revealed a serious problem with the initial core item related to intention-to-treat analysis: when applying the instrument to different trials, four out of the first five reviewers made incorrect assessments for this item (see supplementary appendix 2 for details). This problem

Table 2 | ROBUST-RCT core items and two step approach

Core items and response options	Step 1 Evaluate what happened	Step 2 Judge risk of bias
Core items:		
Item 1 Random sequence generation	Was the allocation sequence adequately generated?	Judge risk of bias related to sequence generation
Item 2 Allocation concealment	Was the allocation adequately concealed?	Judge risk of bias related to allocation concealment
Item 3 Blinding of participants	Were participants blinded?	Judge risk of bias related to blinding of participants
Item 4 Blinding of healthcare providers	Were healthcare providers blinded?	Judge risk of bias related to blinding of healthcare providers
Item 5 Blinding of outcome assessors	Were outcome assessors blinded?	Judge risk of bias related to blinding of outcome assessors
Item 6 Outcome data not included in analysis	Extract the number of participants who were not included in analysis in each group	Judge risk of bias related to the overall percentage of participants not included in analysis
Response options	Definitely yes, probably yes, probably no, definitely no (except for item 6)	Definitely low, probably low, probably high, definitely high

ROBUST-RCT=Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials.

led the panel to drop the intention-to-treat item from the core items list and to modify an existing core item related to missing data to deal with the problem of participants whose outcome data were not included in the analysis for whatever reason (missing outcome data or per protocol analysis), which became the ultimate core item 6 (table 2). In addition, the panel added the failure to avoid an as treated analysis as optional item 6 (table 3). After we made the revision, junior systematic reviewers in the subsequent user tests consistently assessed the core items correctly.

For presentation of the instrument, user testing with systematic review experts revealed that version A (evaluate what happened) might not work well in practice: review experts questioned the rationale for only using version A. One expert suggested combining the two versions into a single instrument with two steps for assessing risk of bias: evaluate what happened and judge risk of bias based on what happened. As regards the two options for instrument presentation (two versions or the two step approach), five review experts expressed their preferences: two opted for the two versions and three opted for the two step approach. The panel ultimately decided to adopt the single instrument with the two step approach while providing the option that systematic reviewers assessing individual trials could choose to complete only step 1—this option incorporates the flexibility and advantage of the two versions approach.

ROBUST-RCT

Supplementary appendix 4 presents ROBUST-RCT. Supplementary appendix 5 provides an Excel sheet in which systematic reviewers can enter their risk of bias assessment and thus generate a risk of bias assessment table for all trials in a systematic review. Supplementary appendix 6 presents the manual with instructions to help systematic review leaders coordinate the risk of bias assessment, and it provides explanations and examples for each item to assist systematic reviewers to complete the instrument. We will provide visualisation and any update about ROBUST-RCT at https://www.clarityresearch.ca/robust-rct.

Core items

Ultimately, ROBUST-RCT included six core items (table 2, also see supplementary appendix 4). Each core item includes two steps for assessing risk of bias. The first step is to evaluate what happened—that is, whether the methodological safeguard had been implemented (eg, step 1 of item 3 judges if participants were blinded). For all but the last item, response options include definitely yes, probably yes, probably no, and definitely no. The second step is to judge risk of bias based on what happened (eg, step 2 of item 3 judges risk of bias related to blinding of participants). The second step requires members of the systematic review team to decide the extent to which any deficits in instituting methodological safeguards resulted in risk of bias. Response options for risk of bias include definitely low, probably low, probably high, and definitely high.

Systematic reviewers assessing individual trials (ie, risk of bias assessors) can complete both steps. However, for core items 1-5, if the risk of bias assessors that the systematic review team recruits are less experienced and may face difficulty in judging risk of

Table 3 | ROBUST-RCT optional items*

Optional items	Titles
Item 1	Whether baseline prognostic factors were balanced between groups
Item 2	Whether co-interventions were balanced between groups in blinded trials
Item 3	Whether outcome assessment or data collection differed between groups
Item 4	Whether follow-up time, frequency, or intensity of outcome assessment differed between groups
Item 5	Whether outcome measurement method was valid (ie, validity of outcome measurement)
Item 6	When investigators conducted an as treated analysis, was the percentage of participants not analysed in the groups to which they were randomised sufficiently low
Item 7	Whether there was selective reporting
Item 8	Whether the trial was terminated early for benefit

ROBUST-RCT=Risk Of Bias instrument for Use in SysTematic reviews-for Randomised Controlled Trials.
*Refer to the manual (see supplementary appendix 6) for considerations when systematic reviewers might or might not include the optional items.

bias, systematic review leaders can ask the risk of bias assessors to complete only step 1 and leave step 2 for the systematic reviewers with more experience.

For core item 6 (outcome data not included in analysis), two approaches can be used to deal with risk of bias. One approach is deciding the risk of bias associated with this item for each individual trial. In this case, systematic review teams will need to set the missing percentage threshold for each response option for step 2 of item 6 (see supplementary appendix 6 for instructions). Risk of bias assessors will determine the percentage of people not included in analysis and where that percentage falls in the risk of bias categories.

An alternative approach for core item 6 involves systematic review teams assessing risk of bias associated with missing data across the entire body of evidence at the meta-analysis level.[19 20] Systematic review teams first need to conduct a complete case analysis; then, to test whether the inference from the complete case analysis is robust, they will perform an analysis imputing data for participants in each trial who were not included in the analysis.

For example, if for a binary outcome the complete case analysis suggests the intervention decreases the risk of an undesirable event, systematic reviewers can conduct a sensitivity analysis assuming, in the control group, the event rate in the participants not included in the analysis is the same as that in the participants who were included in the analysis. Using plausible worse case assumptions, the reviewers can further assume, in the intervention group, the event rate in the participants who were not included in the analysis is higher than that in those included in the analysis.[19 20] Using this approach, risk of bias assessors need complete only step 1 in which they extract the numbers of participants who were not included in the analysis.

Optional items

The instrument includes eight optional items that systematic review teams could consider bringing to the attention of the risk of bias assessors (table 3, also see supplementary appendix 6 for details).

Discussion

We developed ROBUST-RCT, a simply structured and user friendly instrument for assessing risk of bias of randomised controlled trials in systematic reviews. ROBUST-RCT provides six core items, each of which includes two steps: to evaluate what happened in individual trials and to judge the risk of bias based on what happened. ROBUST-RCT also provides eight optional items that systematic reviewers might consider relevant in specific circumstances.

Strengths and limitations of ROBUST-RCT development

We conducted preparatory work to support the development of ROBUST-RCT: a thorough collection of potential candidate items through a survey of existing risk of bias instruments of randomised controlled trials with an assessment of whether the potential items dealt with risk of bias or with other concerns such as indirectness.[5] That process resulted in the items being organised into three categories: assessing risk of bias, assessing concerns other than risk of bias, and possibly assessing risk of bias.[5] A second major aspect of preparatory work was a systematic survey of meta-epidemiological studies that had dealt with the impact of potential risk of bias items on effect estimates in randomised controlled trials.[18] An international panel that reviewed the preparatory material and created the instrument was balanced for both geographical location and sex, and it included experts chosen on the basis of previous publication of risk of bias methodological papers, as well as experienced educators in evidence based medicine.

We developed rigorous criteria for item selection (see box 1) that proved of great use in deciding on the inclusion or exclusion of items (table 1, also see supplementary appendix 3). These criteria measured the items from different dimensions in a comprehensive and clear way: theoretical considerations, empirical support, and two factors—information required to make judgment is commonly reported, and non-expert reviewers can easily assess the item—geared to optimise the practical application of ROBUST-RCT.

User testing involved both junior and experienced senior systematic reviewers. User testing resulted in considerable refinement of the items and presentation of the instrument (see supplementary appendix 2) and ultimately confirmed the simplicity and ease of practical application of ROBUST-RCT: junior systematic reviewers were able to assess the core items correctly (see supplementary appendix 2). However, user testing was limited by the relatively small number of systematic reviewers who participated.

The panellists reached consensus mainly through open discussion rather than more structured approaches, such as the Delphi process. Open discussion was suitable in this case because issues of risk of bias are complex and interconnected. For instance, consideration of whether the co-interventions were balanced between groups involved several questions: what is a sufficiently important co-intervention?; if a co-intervention is sufficiently important, when is imbalance enough to consider as high risk of bias?; is the imbalance in co-interventions a concern of risk of bias or a function of the effect of intervention?; and how easy would it be for non-expert systematic reviewers to make judgments on the aforementioned questions? Ultimately the panel decided that these judgments were too complex for many junior reviewers and should be included in one of the optional rather than core items. The result of the deliberation process was a rich discussion of the relevant considerations.

A limitation of ROBUST-RCT is that it only assesses risk of bias in individually randomised parallel group trials. This weakness will present a challenge for systematic review teams whose review includes cluster or crossover randomised controlled trials. In such a situation, systematic review teams will need to refer

to relevant items in instruments dealing specifically with these study designs. Our group plans to develop extensions of ROBUST-RCT to other trial designs such as cluster or crossover trials.

Relation to previous work

In recent years, the risk of bias assessment process has become overly complex.[9 11] ROBUST-RCT was designed to deal with this problem by focusing on the target users' pragmatic use of the new instrument. Strategies to achieve this goal were item selection criteria that include availability of the information required to make judgments and ease of judgments by junior systematic reviewers (see box 1) and the user testing exercises. No previous instrument had such criteria for item selection, and although Cochrane conducted user testing of RoB 2, challenges in its application suggested that its user testing did not focus on ease of application by junior members of the review team.

We considered the two steps in risk of bias assessment, which in previous instruments are often combined and thus resulted in problematic ambiguity: firstly, to evaluate whether a methodological safeguard had been implemented, and, secondly, to determine whether failure to implement the methodological safeguard resulted in risk of bias. Including two separate steps for assessing these different constructs increases the transparency and conceptual clarity of ROBUST-RCT. Taking into account the different level of experience and expertise across systematic review teams, we offer flexibility about who completes the second step for items 1-5. A review team might require initial risk of bias assessors to complete both evaluations, or require them (if they are less experienced) to complete only step 1 to maximise reliability while leaving the ultimate risk of bias judgment to more experienced systematic review leaders. For item 6, two steps represent two approaches to assessing risk of bias for this item.

Compared to the first Cochrane risk of bias instrument,[6] instructions for ROBUST-RCT offer suggestions about how to classify trials into categories when they fail to report methodological safeguards clearly. This allows reviewers to make reasonable inferences and to classify the trials as probably implemented the methodological safeguards (probably yes in step 1) or probably not implemented the safeguards (probably no in step 1).

Implications

ROBUST-RCT is a new rigorously developed, simply structured, and user friendly instrument for assessing risk of bias of randomised controlled trials in systematic reviews. We believe that ROBUST-RCT has achieved our aim of an optimal balance between simplicity and methodological rigour and is a risk of bias instrument that can be used by review teams with different levels of expertise. Although extensive pretesting of ROBUST-RCT provided evidence of its feasibility and acceptability, wider use may reveal limitations that we can then correct. We therefore encourage future users to bring any limitations to our attention.

AUTHOR AFFILIATIONS

[1]Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China
[2]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, L8S 4L8, Canada
[3]Department of Medicine, Lahey Hospital and Medical Centre, Burlington, MA, USA
[4]Department of Clinical Research, CLEAR-Methods Centre, University Hospital Basel and University of Basel, Basel, Switzerland
[5]Institute for Evidence-Based Healthcare, Bond University, Gold Coast, Queensland, Australia
[6]Department of Anaesthesia, McMaster University, Hamilton, ON, Canada
[7]Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
[8]University of Applied Sciences, Faculty of Business and Social Sciences, Osnabrück, Germany
[9]Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada
[10]Department of Neonatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[11]Liverpool Clinical Trials Clinical Trials Research Centre, Department of Biostatistics, University of Liverpool, Liverpool, UK
[12]Gastro Unit, Copenhagen University Hospital Hvidovre, Copenhagen, Denmark
[13]Department of Statistics, The University of Warwick, Coventry, UK
[14]Department of Internal Medicine, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile
[15]Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), INSERM, Université Paris Descartes, Paris, France
[16]School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[17]York Trials Unit, Department of Health Sciences, University of York, York, UK
[18]Department of Medicine, McMaster University, Hamilton, ON, Canada
[19]MAGIC Evidence Ecosystem Foundation, Oslo, Norway

We thank Rod Jackson for his contribution to development of the instrument. We thank Chunyang Bai, Vamsikalyan Borra, Fatemeh Mirzayeh Fashami, Yasaman Hamidianshirazi, Tanvir Jassal, Sumanth Khadke, Bhargav Makwana, Sahith Rajkumar, Christof Schönenberger, Johannes Schwenke, Jessyca Silva, Michael Wu, Daniel Xie, Shuihua Yang, Puwen Zhang, Qiukui Hao, Susan Hillier, Meixuan Li, Gabriele Meyer, Vassiliki Sinopoulou, Nicole Skoetz, Kehu Yang, and Liang Yao for participating in user testing of the instrument. We thank Prashanti Eachempati for her idea about the Excel sheet including the drop-down listing and colour coding, which inspired the creation of the Excel sheet used in this publication. We thank Stefan Schandelmaier, Prashanti Eachempati, and Derek Chu for their suggestions and comments after using the instrument.

Contributors: GG and YW developed the initial idea. Operations committee members (GG, YW, RBP, RAS, DZ) identified the need for a new instrument, developed a protocol, recruited the panel, organised materials, presented proposals to the panel, and created drafts of the instrument and associated materials. Panel members (GG, YW, RBP, RAS, DZ, MB, PG, EAA, SAO, DB, CG, LLG, JLH, PR, KFS, DJT, SK, RJ, LML) participated in the open discussion in panel meetings. GG and YW are co-chairs of the panel meetings. YW and GG drafted the manuscript, and others reviewed and revised the manuscript. All authors approved the final version of the manuscript, instrument, and manual. GG and YW are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This work was supported by the Einstein Foundation Berlin as part of the Einstein Foundation Award for Promoting

Quality in Research. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, the Einstein Foundation or the award jury. The funder had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: support from the Einstein Foundation Berlin; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The Hamilton Integrated Research Ethics Board approved the user testing study.

Patient and public involvement: No patients or members of the public were involved in this work.

Dissemination to participants and related patient and public communities: We sent the user testing results, revision we made based on user testing, and latest version of the instrument to the participants of user testing. To disseminate the new instrument to related public communities, we previewed the new instrument on X (formerly Twitter) and provided contact information for access to it.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Guyatt GH, Rennie D, Meade MO, et al. *Users' guide to the medical literature: a manual for evidence-based clinical practice.* McGraw-Hill Education, 2015.
2 Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001;54:651-4. doi:10.1016/S0895-4356(00)00360-7
3 Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383-94. doi:10.1016/j.jclinepi.2010.04.026
4 Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407-15. doi:10.1016/j.jclinepi.2010.07.017
5 Wang Y, Ghadimi M, Wang Q, et al. Instruments assessing risk of bias of randomized trials frequently included items that are not addressing risk of bias issues. *J Clin Epidemiol* 2022;152:218-25. doi:10.1016/j.jclinepi.2022.10.018
6 Higgins JP, Altman DG, Gøtzsche PC, et al, Cochrane Bias Methods GroupCochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928. doi:10.1136/bmj.d5928
7 Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol* 2012;65:262-7. doi:10.1016/j.jclinepi.2011.04.015
8 Savović J, Weeks L, Sterne JA, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. *Syst Rev* 2014;3:37. doi:10.1186/2046-4053-3-37
9 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898
10 Moore THM, Higgins JPT, Dwan K. Ten tips for successful assessment of risk of bias in randomized trials using the RoB 2 tool: Early lessons from Cochrane. *Cochrane Evidence Synthesis and Methods* 2023;1:e12031doi:10.1002/cesm.12031.
11 Kuehn R, Wang Y, Guyatt G. Overly complex methods may impair pragmatic use of core evidence-based medicine principles. *BMJ Evid Based Med* 2024;29:139-41. doi:10.1136/bmjebm-2024-112868
12 Babić A, Barcot O, Visković T, et al. Frequency of use and adequacy of Cochrane risk of bias tool 2 in non-Cochrane systematic reviews published in 2020: Meta-research study. *Res Synth Methods* 2024;15:430-40. doi:10.1002/jrsm.1695
13 Martimbianco ALC, Sá KMM, Santos GM, Santos EM, Pacheco RL, Riera R. Most Cochrane systematic reviews and protocols did not adhere to the Cochrane's risk of bias 2.0 tool. *Rev Assoc Med Bras (1992)* 2023;69:469-72. doi:10.1590/1806-9282.20221593
14 Crocker TF, Lam N, Jordão M, et al. 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. *J Clin Epidemiol* 2023;161:39-45. doi:10.1016/j.jclinepi.2023.06.015
15 Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. *J Clin Epidemiol* 2020;126:37-44. doi:10.1016/j.jclinepi.2020.06.015
16 Babic A, Pijuk A, Brázdilová L, et al. The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey. *BMC Med Res Methodol* 2019;19:77. doi:10.1186/s12874-019-0718-8
17 Jørgensen L, Paludan-Müller AS, Laursen DR, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. *Syst Rev* 2016;5:80. doi:10.1186/s13643-016-0259-8
18 Wang Y, Parpia S, Couban R, et al. Compelling evidence from meta-epidemiological studies demonstrates overestimation of effects in randomized trials that fail to optimize randomization and blind patients and outcome assessors. *J Clin Epidemiol* 2024;165:111211. doi:10.1016/j.jclinepi.2023.11.001
19 Goldkuhle M, Bender R, Akl EA, et al, GRADE Working Group. GRADE Guidelines: 29. Rating the certainty in time-to-event outcomes-Study limitations due to censoring of participants with missing data in intervention studies. *J Clin Epidemiol* 2021;129:126-37. doi:10.1016/j.jclinepi.2020.09.017
20 Guyatt GH, Ebrahim S, Alonso-Coello P, et al. GRADE guidelines 17: assessing the risk of bias associated with missing participant outcome data in a body of evidence. *J Clin Epidemiol* 2017;87:14-22. doi:10.1016/j.jclinepi.2017.05.005

Supplementary information: Appendix 1—Protocol for development of ROBUST-RCT
Supplementary information: Appendix 2—Details of user testing
Supplementary information: Appendix 3—Panel's initial decision on item selection and instrument
Supplementary information: Appendix 4—Details of ROBUST-RCT in pdf format
Supplementary information: Appendix 5—Excel spreadsheet
Supplementary information: Appendix 6—Manual for ROBUST-RCT