Journal Club
02/05/2025

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

Nattapat Khumtong

6636457

RESEARCH METHODS AND REPORTING

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

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]

Outline

- Introduction
- Historical Background: RoB and RoB 2
- Why Develop ROBUST-RCT?
- Development Process of ROBUST-RCT
- Results
- Strengths and Limitations of ROBUST-RCT

Introduction

- Systematic reviews of randomised controlled trials provide the best evidence for the effects of healthcare interventions

- Flaws in trial design and conduct may result in biased estimates of effects and misleaded conclusions.

- Risk of bias assessment of randomised controlled trials is an essential step in the systematic review process.

Previous risk of bias assessment tools

Risk of Bias assessment tool (RoB)

- The first Cochrane risk of bias instrument by the Cochrane Collaboration.
- Introduced in 2008
- Included an “unclear” response option >> failed to take advantage of reasonable inferences about the presence or absence of risk of bias.
- Users have reported problems with assessing the incomplete outcome data and the selective reporting domains.

Previous risk of bias assessment tools

The revised Cochrane instrument for assessing risk of bias in randomised controlled trials (RoB 2)

- Revised in 2019
- Complex algorithms (up to 7 signaling questions/domain).
- Terminology difficulties (e.g., "deviations from intended intervention").
- Poor uptake outside Cochrane; frequent misapplication.
- Low interrater reliability even among experienced users

Example

Risk of bias assessment

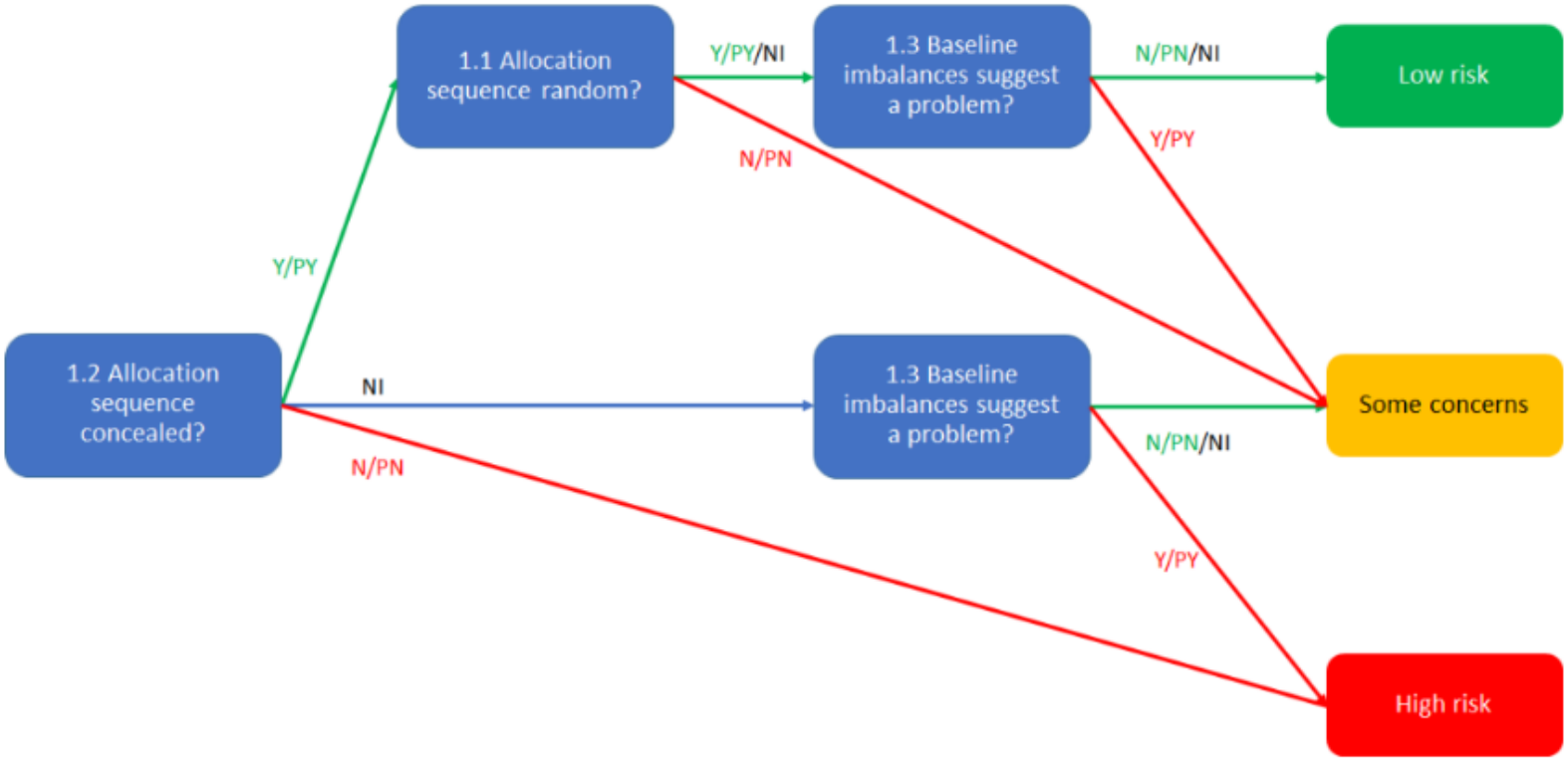
Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Domain 1: Risk of bias arising from the randomization process

Signalling questions	Comments	Response options
1.1 Was the allocation sequence random?		Y / PY / PN / N / NI
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?		Y / PY / PN / N / NI
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?		Y / PY / PN / N / NI
Risk-of-bias judgement		Low / High / Some concerns
Optional: What is the predicted direction of bias arising from the randomization process?		NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

Signalling questions	Comments	Response options
2.1. Were participants aware of their assigned intervention during the trial?		Y / PY / PN / N / NI
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?		Y / PY / PN / N / NI
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?		NA / Y / PY / PN / N / NI
2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?		NA / Y / PY / PN / N / NI
2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?		NA / Y / PY / PN / N / NI
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?		Y / PY / PN / N / NI
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?		NA / Y / PY / PN / N / NI
Risk-of-bias judgement		Low / High / Some concerns
Optional: What is the predicted direction of bias due to deviations from intended interventions?		NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable



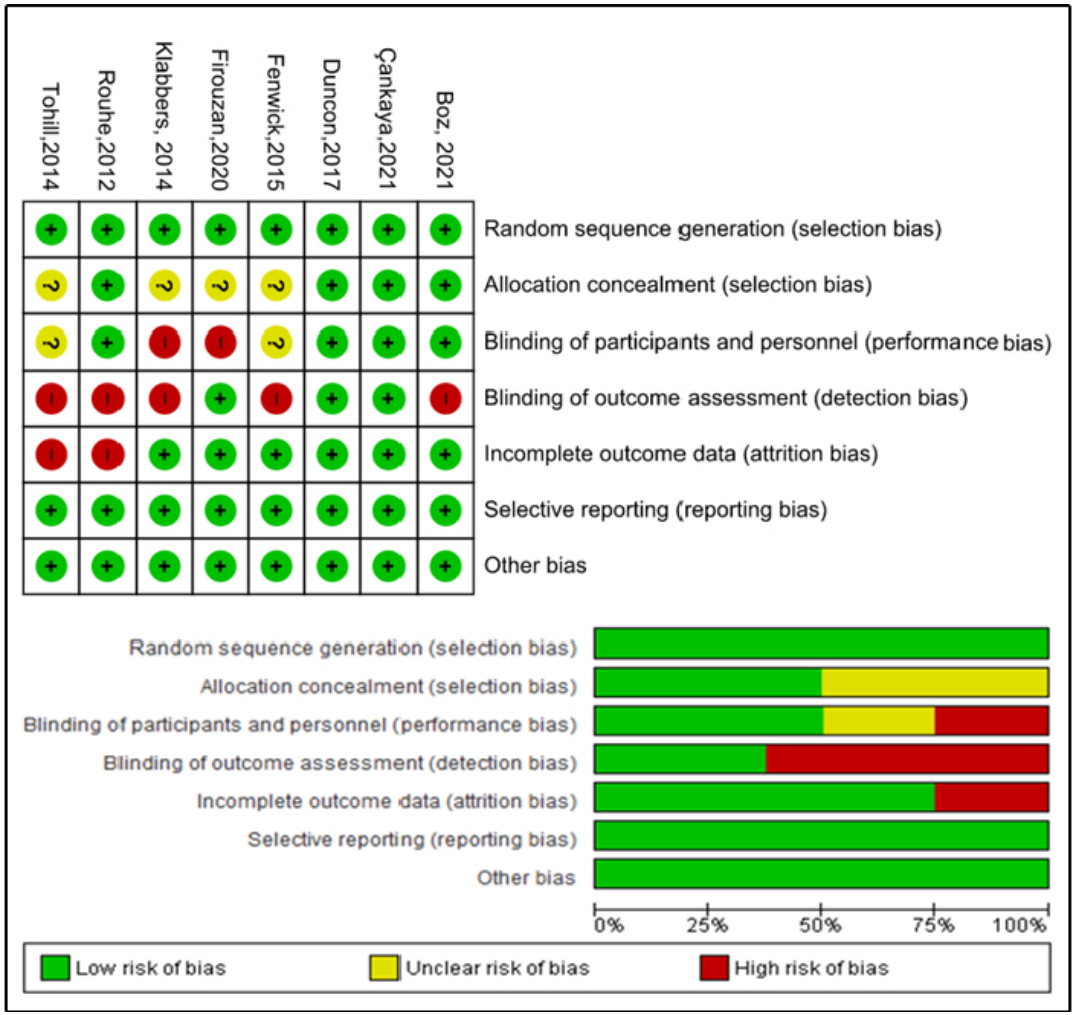
Algorithm for suggested judgement of risk of bias arising from the randomization process

Overall risk-of-bias judgement

- **Low risk of bias >> All domains for low risk of bias.**

- Some concerns >> Some concerns in at least one domain, but not to be at high risk of bias for any domain.

- High risk of bias >> At least one domain for high risk of bias. Or
There are some concerns for multiple domains in a way that substantially lowers confidence in the result.


Tohill,2014
Rouhe,2012
Klabbers, 2014
Firouzan,2020
Fenwick,2015
Duncon,2017
Çankaya,2021
Boz, 2021
Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
0%
25%
50%
75%
100%
Low risk of bias
Unclear risk of bias
High risk of bias

Why Develop ROBUST-RCT?

RoB2 challenges:

- Too sophisticated for many users.
- Difficult for less-experienced systematic reviewers.

A new tool needed:

- **Balance** between simplicity and rigor.
- **Ease of use** for junior and senior team members alike.
- Concept: practical without sacrificing scientific quality.

Development Process of ROBUST-RCT : Who ?

The instrument development team

1. Operations committee

- Members of the operations committee (GG, YW, RBP, RAS, DZ)
- Identified the need for a new instrument
- Developed a protocol
- Recruited the panel of experts
- Presented proposals to the panel
- Created drafts of the instrument and associated materials.

Development Process of ROBUST-RCT : Who ?

2. Panel

- The operations committee identified experts in risk of bias assessment from the author lists of methodological papers >> 295 eligible papers
- Panel membership had to 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, 10 were randomly selected >> **9 agreed.**
- included **2 more methodological experts (MB, PG)** who the committee members knew and thought could make substantial contributions.
- Included **3 experienced educators in evidence based medicine** (SK, RJ, LML)
- **Total 19 members** (International collaboration included 10 men and 9 women)

Development Process of ROBUST-RCT : Ground rules for instrument development

- Aims to **assess risk of bias of RCTs in the context of systematic reviews.**
- **User friendly** instrument
- Bias defines as a systematic error or systematic deviation from the truth.
- 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.**
- **For individually randomized parallel group trials.** Not for cluster trials and crossover trials is for future consideration.
- This instrument will not include items for the detection of fraud.

Development Process of ROBUST-RCT : Collection of candidate items

17 risk of bias instruments of randomized controlled trials published from 2010 to October 2021

Item classification by 13 panellists

Category 1
Majority of the panellists judged as addressing risk of bias

Category 2
Majority of the panellists judged as not addressing risk of bias

Category 3
Substantial disagreement among the panellists

Development Process of ROBUST-RCT : How to select the items

- Empirical evidence from meta-epidemiological studies for item selection >> examining the impact of potential risk of bias problems (items in categories 1 and 3) on effect estimates in RCTs.

- E.g. " Inadequate random sequence generation and allocation concealment probably lead to effect overestimation"

Development Process of ROBUST-RCT : How to select the items

Category 1
judged as addressing risk of bias

Category 2
judged as not addressing risk of bias

Category 3
Substantial disagreement

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 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

Development Process of ROBUST-RCT : How to select the items

- 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 and **optional items** for the instrument.

Development Process of ROBUST-RCT : User testing for junior reviewers

- Enrolled 15 people who had assessed risk of bias in RCTs for at least one systematic review and had never led any systematic review of RCTs
- 5 trials that presented challenges in risk of bias assessment were selected by the panelists
- 2 committee members (YW and GG) assessed risk of bias in these trials
- Each participant received one trial, the draft of the instrument, and the manual.
- YW conducted a think-aloud interview of 1 hour with each participant.
- YW compared the participant's assessment with the assessment made and agreed on by YW and GG
- Participants expressed their overall experience in applying the instrument.

Table 1. Characteristics of the junior systematic reviewers in the user-testing

Characteristic	Number of participants (total 15)
Female	7
Country	
Canada	7
China	3
India	2
Switzerland	2
US	1
Clinical background	
Physician	7
Pharmacist	2
Dietitian	1
No clinical background	5
Student status	
PhD student	2
Master student	6
Undergraduate student	3
Not student	4
Number of systematic reviews in which they have assessed risk of bias of randomized trials	
1-2	10
3-5	3
>5	2

Table 2. Trials that used in user testing with junior systematic reviewers

1. Chochinov HM, Kristjanson LJ, Breitbart W, et al. Effect of dignity therapy on distress and end-of-life experience in terminally ill patients: a randomised controlled trial. Lancet Oncol. 2011 Aug;12(8):753-62. doi: 10.1016/S1470-2045(11)70153-X.
2. Hansen CD, Gram-Kampmann EM, Hansen JK, et al. Effect of Calorie-Unrestricted Low-Carbohydrate, High-Fat Diet Versus High-Carbohydrate, Low-Fat Diet on Type 2 Diabetes and Nonalcoholic Fatty Liver Disease : A Randomized Controlled Trial. Ann Intern Med. 2023 Jan;176(1):10-21. doi: 10.7326/M22-1787.
3. Lou W, Xia Y, Xiang P, et al. Prevention of upper gastrointestinal bleeding in critically ill Chinese patients: a randomized, double-blind study evaluating esomeprazole and cimetidine. Curr Med Res Opin. 2018 Aug;34(8):1449-1455. doi: 10.1080/03007995.2018.1464132.
4. Nayyab I, Ghous M, Shakil Ur Rehman S, et al. The effects of an exercise programme for core muscle strengthening in patients with low back pain after Caesarian-section: A single blind randomized controlled trial. J Pak Med Assoc. 2021 May;71(5):1319-1325. doi: 10.47391/JPMA.596.
5. Parekh DJ, Reis IM, Castle EP, et al. Robot-assisted radical cystectomy versus open radical cystectomy in patients with bladder cancer (RAZOR): an open-label, randomised, phase 3, non-inferiority trial. Lancet. 2018 Jun 23;391(10139):2525-2536. doi: 10.1016/S0140-6736(18)30996-6.

Development Process of ROBUST-RCT : User testing for experts reviewers

- Searched the Cochrane Library 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 RCTs** (not limited to Cochrane reviews) will be invited

- **8 participants** received the instrument and manual.
- YW followed a semistructured interview guide, interviewing each participant for 1 hour and transcribed the interviews.
- Identified concerns and solutions and presented these to the panel in deciding on modifications to the instrument and manual.

Result : Panel's initial decision

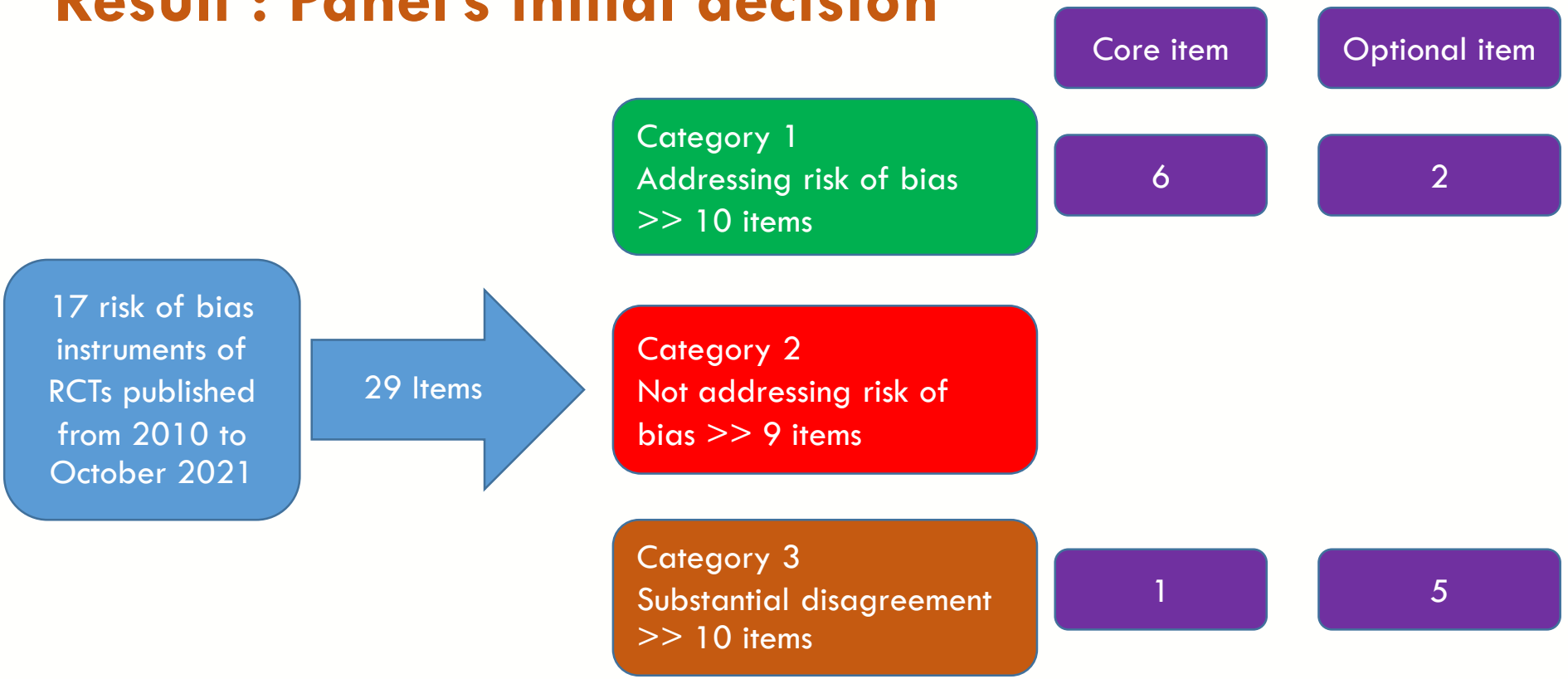


Table 1 | Initially selected core items and optional items and judgment about whether they met the six criteria for item selection*

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)

Result : Panel's initial decision

Table 2. Decisions for items in category 2 (items that the majority judged as not addressing risk of bias in a survey)*

Items	Item selection decision	Reasons for excluding
Whether the outcome measurement was reliable (i.e., reliability of outcome measurement)	Exclude	Imprecision (random error) rather than risk of bias.
Whether the follow-up time was adequate to identify the outcome of interest	Exclude	Indirectness (applicability) rather than risk of bias.
Whether the sample size was big enough	Exclude	Imprecision (random error) rather than risk of bias.
Whether the sampling approach (approach to selecting participants from whole population) was appropriate	Exclude	Indirectness (applicability) rather than risk of bias.
Whether there was conflict of interest	Exclude	Not directly related to risk of bias – may influence effect estimate but must through other mechanisms.
Whether there was funding	Exclude	Not directly related to risk of bias – may influence effect estimate but must through other mechanisms.
Whether the results were comparable for all trial sites	Exclude	Not related to risk of bias.
Whether there was run-in period before randomization	Exclude	Not related to risk of bias.
Whether the inclusion and exclusion criteria were appropriate	Exclude	Indirectness (applicability) rather than risk of bias.

Result : Revision based on user testing

Core items with 2 steps approach

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

Result : Revision based on user testing

Core items with 2 steps approach

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

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

First : evaluate what happened, whether the methodological safeguard had been implemented

Result : Revision based on user testing

Core items with 2 steps approach

Core items and response options	Step 2 Judge risk of bias
Core items:	
Item 1 Random sequence generation	Judge risk of bias related to sequence generation
Item 2 Allocation concealment	Judge risk of bias related to allocation concealment
Item 3 Blinding of participants	Judge risk of bias related to blinding of participants
Item 4 Blinding of healthcare providers	Judge risk of bias related to blinding of healthcare providers
Item 5 Blinding of outcome assessors	Judge risk of bias related to blinding of outcome assessors
Item 6 Outcome data not included in analysis	Judge risk of bias related to the overall percentage of participants not included in analysis
Response options	Definitely low, probably low, probably high, definitely high

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.

Result : Revision based on user testing

Core items with 2 steps approach

Item 1 Random sequence generation

Step 1: Was the allocation sequence adequately generated
Step 2: Judge risk of bias related to sequence generation
Instructions for step 1 and step 2 are the same.

Definitely Yes/Low
Trial explicitly stated use of an adequate method of generating the random allocation sequence. Examples include: random number table; random number generator; throwing dice; drawing of lots; minimization.
Explanation:
Adequate method of generating the random allocation sequence refers to the method that incorporates a random element and thus can generate a random and unpredictable sequence. *Minimization* is a method of ensuring intervention groups are closely similar for multiple prognostic factors, even in small trials.[3] Using minimization, the first participant is allocated randomly. For each subsequent participant, investigators determine assignment to the intervention that would lead to better balance between the groups over all the identified prognostic factors.
Example: A trial stated "Randomization was performed with a computer-generated allocation sequence...".[4]

Result : Revision based on user testing

Optional items

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

Table 3. Optional items

	Reasons for not including as a core item	Reasons for including as an optional item in this instrument	Considerations regarding inclusion of this item in a systematic review
Optional item 1: **Whether baseline prognostic factors were balanced between groups**	• When considering this item, one has to consider whether a baseline characteristic is an important prognostic factor in the particular context. • We have already included random sequence generation and allocation concealment as core items. Prognostic imbalance due to inappropriate randomization will be at least in part covered by these core items. If randomization is conducted properly and sample size is sufficient, prognostic imbalance would happen rarely and be simply due to chance. • While prognostic imbalance will often happen in small studies, it is much less likely across the entire range of studies (prognostic imbalance in one study is likely to be ameliorated by distribution of prognostic variables in other studies). • Empirical evidence regarding this item actually creating bias is very uncertain.	• When sample size is small, investigators may generate sequence appropriately and conceal and still have imbalance of prognostic factors simply by chance. • When a baseline characteristic with known appreciable prognostic power is imbalanced between groups, it may create serious bias.	• If there is problem with random sequence generation or allocation concealment (thus high risk of bias related to sequence generation or concealment item), no need to include this item. • If random sequence generation and allocation concealment performed well, and there is an important imbalance in important prognostic factors, an extra risk of bias exists. This item captures this problem. • The larger the imbalance in a factor, the stronger the prognostic power of the factor, the larger the number of trials in which this exists and the larger the weight of these trials in the meta-analysis, the more likely one would include this item.

Discussion

- ROBUST-RCT provides 6 core items, each of which includes two steps: to evaluate what happened in individual trials and to judge the risk of bias

- 8 optional items that systematic reviewers might consider relevant in specific circumstances.

Strengths

- Preparatory development work including collection of potential candidate items through a survey of existing risk of bias instruments and systematic survey of meta-epidemiological studies
- Open discussion, suitable in this case because issues of risk of bias are complex and interconnected
- Simplicity and ease of practical application of ROBUST-RCT

Limitations

- Only assesses risk of bias in individually randomised parallel group trials.

Comparison with RoB2

	ROBUST-RCT	RoB2
Item to be answers	6 core items 8 optional items	5 domains 3-7 items per domain
2-step approach	Yes	Mixed
Complexity	Moderate	High
User-friendliness	High	Low
Focus on risk of bias problem (not imprecision, publication bias, or reporting quality)	Yes	No

Thank you

For your attention

Question and answer