Abstract

JBI have recently began the process of updating and revising their suite of critical appraisal tools to ensure that these tools remain compatible with recent developments made within risk of bias science. Following a rigorous development process led by the JBI Effectiveness Methodology Group, this paper presents the revised critical appraisal tool for the assessment of risk of bias of randomized controlled trials.

This paper also presents practical guidance on how the questions of this tool are to be interpreted and applied by systematic reviewers, while providing topical examples. We also discuss the major changes made to this tool compared to the previous version and justification for why these changes facilitate best-practice methodologies in this field.

Introduction

Systematic reviews are a foundational and fundamental component in the practice of evidence-based health care. Systematic reviews involve the collation and synthesis of the results of multiple independent studies that address the same research question. Prior to the creation of these synthesized results, all studies that have been selected for inclusion (i.e. those that meet the *a priori* eligibility criteria)[1] need to be subjected to a process of critical appraisal.[2,3] The purpose of this appraisal (for quantitative evidence) is to determine the extent to which a study has addressed the possibility of bias in its design, conduct and analysis. By subjecting every study included in a systematic review to rigorous critical appraisal, it allows reviewers to appropriately consider how the conduct of individual studies may impact the synthesized result, thus enabling the synthesized result to be correctly interpreted.[4]

Recent advancements in the science of risk of bias assessment[5-7] have argued that only questions related to the internal validity of that study, should be considered in the assessment of that study's inherent biases. The assessment of a study's risk of bias often occurs during a structured and transparent critical appraisal process. For example, a question related to how generalizable a participant sample is to the broader population does not impact on that study's internal validity, and thus it's inherent biases,[5-8] but is still useful to describe the external validity of that study. There is also now an expectation, that assessments of bias occur at different levels, including outcome level and result level assessments, that may be different within the same study depending on the outcome or result being assessed.[5,8] These (and other) advancements have been discussed previously in an introduction to this body of work.[8]

It is acknowledged that the existing suite of JBI critical appraisal instruments are not aligned to these recent advancements and conflate and confuse the process of 'critical appraisal' with that of 'risk of bias assessment'. Therefore, the JBI Effectiveness Methodology Group, under the auspices of the JBI Scientific Committee updated the entire suite of JBI critical appraisal tools to be better aligned to best practice methodologies.[8] This paper serves to introduce the revised critical appraisal tool for randomized controlled trials and to provide step-by-step guidance for how to use and implement this tool in future systematic reviews. We also clearly document and justify each major change made for this revised tool.

Methods

In 2021, a working group of researchers and methodologists known as the JBI Effectiveness Methodology Group was tasked by the JBI Scientific Committee[9] to revise the current suite of

JBI critical appraisal tools for quantitative analytical study designs. The aim of this work was to improve the longevity and usefulness of these tools and to reflect current advancements made in this space,[5-7] whilst adhering to the reporting and methodological requirements as established by PRISMA 2020[10] and GRADE.[11] To summarise this process, the JBI Effectiveness Methodology Group began with cataloguing the questions asked in each JBI critical appraisal tool for study designs that employ quantitative data. These questions were ordered into constructs of validity (internal, statistical conclusion, comprehensiveness of reporting, external) through a series of roundtable discussions between members of the JBI Effectiveness Methodology Group. For questions that were related to the internal validity construct, they were further catalogued to a domain of bias through a series of mapping exercises and round-table discussions. Finally, questions were then separated based on whether they were answered at either the study, outcome, or result level. The full methodological processes undertaken for this revision, including the rationale for all decisions made have been documented previously.[8]

How to use the revised tool

The key changes

Similar to previous versions of these tools, the revised JBI critical appraisal tool for RCTs presents a series of questions. These questions aim to identify whether certain safeguards have been implemented by the study to minimize risk of bias or to address other aspects relating to the validity or quality of the study. Each question can be scored as being 'met' (yes), 'unmet' (no), unclear or not applicable. As described previously[8] the wording of these questions presented in the revised JBI critical appraisal tool for RCTs have not been altered in any way from the wording of the questions presented in the previous version of the JBI critical appraisal tool for RCTs.[4] However, the organization of these questions, the order in which they should be addressed and answered, and the means to answer them have been changed.

The questions of this revised tool have been presented according to the construct of validity to which they pertain. The specific validity constructs that are pertinent to the revised JBI critical appraisal tool for RCTs include internal validity and statistical conclusion validity. Questions that have been organized under the 'internal validity' construct have been further organized according to the domain of bias that they are specifically addressing. The domains of bias relevant to the revised JBI critical appraisal tool for RCTs includes bias related to: "selection and allocation", "administration of intervention/exposure", "assessment, detection and measurement of the outcome" and "participant retention". A detailed description of these validity constructs and domains of biases has been reported previously.[8]

The principal difference of the revised JBI critical appraisal tool for RCTs in comparison to its predecessor resides in its structure and organisation, which is now deliberately designed to facilitate judgments related to risk of bias at different levels (e.g. bias at the study level, outcome level or bias at the result level) where appropriate.[8] For the questions that are to be answered at the outcome level (questions 7-12), the tool provides the ability to respond to the questions for up to seven different outcomes. The limit of seven outcomes ensure it aligns with the maximum number of outcomes recommended to be included in a GRADE Summary of Findings Table and/or Evidence Profile .[12] For the questions to be answered at the result level (questions 10-12) the tool presents the option to record a different decision, for three results for each outcome presented (by default). Reviewers may face cases where the number of

outcomes being appraised from a particular RCT are less than seven, and the results being appraised are greater than three per outcome. The tool can be edited as required by the review team to facilitate their use in these cases.

For example, let us consider a hypothetical RCT that has included two outcomes relevant to a question of a systematic review team. These outcomes are mortality and quality of life, both of which have been measured at two time points within the study. When using this tool, questions 1-6 and 13 are universal to both outcomes as they are addressed at the study level and are only answered once. The reviewer should then address question 7-9 twice, once for each outcome that is being appraised. Likewise, question 10-12 should be addressed separately for both outcomes but should also be assessed for each of the results that has contributed data towards that outcome (e.g., mortality at time point 1 and 2). In this example, the reviewer would assess question 10-12 four different times. It is also important to note that, as with other critical appraisal tools[3,13] this tool should also be applied in duplicate and independently during the systematic review process. Reviewers should also be wary to only appraise outcomes that are relevant to their systematic review question. If the only relevant outcome from this RCT for the systematic review question was mortality, then appraising the outcome 'quality of life' would not be expected.

Interpretation of critical appraisal

Some reviewers may take the approach to 'remove' studies from progressing to data extraction or synthesis in their review following this critical appraisal process. Removal of a study following critical appraisal may involve considering whether a certain criterion had not been met (e.g., randomization not being demonstrated may warrant removal, assuming the review was not also including other study designs with lesser internal validity due to not attempting randomization). Another procedure may include the review team weighting each question of the tool (e.g., randomization may be twice as important as blinding of the outcome assessors), if a study fails to meet a predetermined 'weight' (decided by the review team) then it may be removed. Other approaches may be to use simple 'cut-off' scores (e.g. if a study is scored with ten "yes" responses then it is included), or to exclude studies that have been "judged" as having a high-risk of bias.[8] However, we do not recommend that studies are removed from a systematic review following critical appraisal.

By removing studies, it presupposes that the purpose of a systematic review is to only permit 'high-quality' studies being synthesized together. While it may readily promote alignment to the 'best available' evidence, it limits the full potential of the processes of evidence synthesis to fully investigate eligible studies, their data, and provide a complete 'view' of the evidence available to inform the review question.[14,15] There are several other approaches to incorporate the results of critical appraisal into the systematic review or meta-analysis. These approaches can include meta-regression, elicitation of expert opinion, using prior distributions, and quality-effect modelling.[16] However, these techniques demand appropriate statistical expertise, and fall beyond the scope of this paper. Regardless of the approach ultimately decided upon by the reviewers, importantly, the results of the critical appraisal process should always be considered in the analysis and interpretation of the findings of the synthesis.

Overall assessment and presentation of results

Previous iterations of the JBI critical appraisal tool for RCTs intuitively supported reviewers assessing the overall quality of study using a checklist- or scale-based tool structure (each item can be quantified, which is enumerated to provide an overall quality score).[8] The revised tool has been designed to also facilitate judgments specific to the domains of bias in which the questions belong. To provide an example, assuming we have two included studies, their response to questions one, two and three may appear as follows (Table 1):

<insert Table 1 here>

A reviewer may determine that for study 1, there was a low risk of bias for the domain of 'selection and allocation' as all questions responded with a 'yes'. However, for study 2, a reviewer may determine a moderate risk of bias for the same domain, as the response to one of the questions was a 'no'. Importantly, we provide no thresholds for grading of bias severity (i.e. low, moderate, high, critical or other approaches) and leave this to discretion of the user and specific context in which they are working. Considering the questions and assessments in this regard, looking across all included studies (or a single study) can permit the reviewer to readily comment on how the risk of bias may impact on the certainty of their results at this domain-level in the GRADE approach. A judgments-approach as described above is one way for users to adopt the revised JBI critical appraisal tool for RCTs, however the tool is still compatible with either a checklist- or scale-based structure[8] and the decision for which approach to follow is left to the discretion of the review team.

Current tools to appraise RCTs[5] ask the reviewer to establish an overall assessment to the risk of bias for each appraised study and for the overall body of evidence (i.e., all appraised studies). The revised JBI critical appraisal tool for RCTs does not strictly prescribe to this, regardless of the approach followed. However, if reviewers opt to establish an overall assessment, then these assessments should not take into consideration the questions regarding statistical conclusion validity (questions 11,12, and 13). As risk of bias is only impacted by the internal validity construct.[8]

Irrespective of the approach taken, the results of critical appraisal using the revised JBI critical appraisal tool for RCTs should be reported narratively in the review. This narrative summary should provide both an overall description of the methodological quality and risk of bias at the domain level of the included studies. There should also be a statement made regarding any important or interesting deviations from the observed trends. This narrative summary can be supported with the use of a table or graphic that showcases how each included study responded. We recommend presenting the results of critical appraisal for all questions via a table rather than summarizing with a score. For example, a reviewer may use the same example introduced earlier (Table 1), or an additional example (Table 2) attached in the appendix. (Note: these designs are not prescriptive, and only serve as an example).

The Revised JBI Critical Appraisal Tool for Randomized Controlled Trials

The criteria and considerations that should be made by reviewers when answering the questions of the revised JBI critical appraisal tool for RCTs has been provided as Table 3 in the Appendix.

Question 1: Was true randomization used for assignment of participants to treatment groups?

Category: Internal validity

Domain: Bias related to selection and allocation

Appraisal: Study level

If participants are not allocated to treatment and control groups by random assignment there is a risk that this assignment to groups can be influenced by the known characteristics of the participants themselves. These known characteristics of the participants may distort the comparability of the groups (i.e. does the intervention group contain more people over the age of 65 as compared to the control?). A true random assignment of participants to the groups means that a procedure is used that allocates the participants to groups purely based on chance, not influenced by any known characteristics of the participants. Reviewers should check the details about the randomization procedure used for allocation of the participants to study groups. Was a true chance (random) procedure used? For example, was a list of random numbers used? Was a computer-generated list of random numbers used? Was a statistician, external to the research team consulted for the randomization sequence generation? Additionally, reviewers should check that the authors are not stating they have used random approaches when they have instead used systematic approaches (such as allocating by days of the week).

Question 2: Was allocation to groups concealed?

Category: Internal validity

Domain: Bias related to selection and allocation

Appraisal: Study level

If those allocating participants to the compared groups are aware of which group is next in the allocation process, (i.e., the treatment or control group) there is a risk that they may deliberately and purposefully intervene in the allocation of patients. This may result in the preferential allocation of patients to the treatment group or to the control group. This may directly distort the results of the study, as participants no longer have an equal and random chance to belong to each group compared. Concealment of allocation refers to procedures that prevent those allocating patients from knowing before allocation which treatment or control is next in the allocation process. Reviewers should check the details about the procedure used for allocation concealment. Was an appropriate allocation concealment procedure used? For example, was central randomization used? Were sequentially numbered, opaque and sealed envelopes used? Were coded drug packs used?

Question 3: Were treatment groups similar at the baseline?

Category: Internal validity

Domain: Bias related to selection and allocation

Appraisal: Study level

As with question 1, any differences between the known characteristics of participants included in compared groups constitutes a threat to internal validity. If differences in these characteristics do exist, then there is potential that the 'effect' cannot be attributed to the potential 'cause' (the examined intervention or treatment). This is because the 'effect' may be explained by the differences between participant characteristics and not due to the intervention/treatment of interest. Reviewers should check the characteristics reported for participants. Are the participants from the compared groups similar with regards to the characteristics that may explain the effect even in the absence of the 'cause', for example, age, severity of the disease, stage of the disease, co-existing conditions and so on? Reviewers should check the proportions of participants with specific relevant characteristics in the compared groups. [Note: **Do NOT** only consider the P-value for the statistical testing of the differences between groups with regards to the baseline characteristics.]

Question 4: Were participants blind to treatment assignment?

Category: Internal validity

Domain: Bias related to administration of intervention/exposure

Appraisal: Study level

Participants that are aware of their allocation to either the treatment or the control may behave, respond, or react differently to their assigned treatment (or control) than compared to participants that remain unaware of their allocation. Blinding of participants is a technique used to minimize this risk. Blinding refers to procedures that prevent participants from knowing which group they are allocated. If blinding has been followed, participants are not aware if they are in the group receiving the treatment of interest or if they are in any other group receiving the control interventions. Reviewers should check the details reported in the article about the blinding of participants with regards to treatment assignment. Was an appropriate blinding procedure used? For example, were identical capsules or syringes used? Were identical devices used? Be aware of different terms used, blinding is sometimes also called masking.

Question 5: Were those delivering the treatment blind to treatment assignment?

Category: Internal validity

Domain: Bias related to administration of intervention/exposure

Appraisal: Study level

Like question 4, those delivering the treatment that are aware of participant allocation to either treatment or control, may treat participants differently than compared to those that remain unaware of participant allocation. There is the risk that any potential change in behaviour may influence the implementation of the compared treatments and the results of the study may be distorted. Blinding of those delivering treatment is used to minimize this risk. When this level of blinding has been achieved, those delivering the treatment are not aware if they are treating the group receiving the treatment of interest or if they are treating any other group receiving the control interventions. Reviewers should check the details reported in the article about the blinding of those delivering treatment with regards to treatment assignment. Is there any information in the article about those delivering the treatment? Were those delivering the treatment unaware of the assignments of participants to the compared groups?

Question 6: Were treatment groups treated identically other than the intervention of interest?

Category: Internal validity

Domain: Bias related to administration of intervention/exposure

Appraisal: Study level

To attribute the 'effect' to the 'cause', (assuming no bias related to selection and allocation) there should be no other difference between the groups in terms of treatment or care received, other than the treatment or intervention controlled by the researchers. If there are other exposures or treatments occurring at the same time with the 'cause' (the treatment or intervention of interest), then the 'effect' can potentially not be attributed to the examined 'cause' (the investigated treatment). This is because it is plausible that the 'effect' may be explained by these other exposures or treatments that occurred at the same time with the 'cause'. Reviewers should check the reported exposures or interventions received by the compared groups. Are there other exposures or treatments occurring at the same time with the 'cause'? Is it plausible that the 'effect' may be explained by other exposures or treatments occurring at the same time with the 'cause'? Is it clear that there is no other difference between the groups in terms of treatment or care received, other than the treatment or intervention of interest?

Question 7: Were outcome assessors blind to treatment assignment?

Category: Internal validity

Domain: Bias related to assessment, detection and measurement of the outcome

Appraisal: Outcome level

Like question 4 and 5, those assessing the outcomes that are aware of participant allocation to either treatment or control, may treat participants differently than compared to those that remain unaware of participant allocation. Therefore, there is a risk that the measurement of the outcomes between groups may be distorted, and the results of the study may themselves be distorted. Blinding of outcomes assessors is used in order to minimize this risk. Reviewers should check the details reported in the article about the blinding of outcomes assessors with regards to treatment assignment. Is there any information in the article about outcomes assessors? Were those assessing the treatment's effects on outcomes unaware of the assignments of participants to the compared groups?

Question 8: Were outcomes measured in the same way for treatment groups?

Category: Internal validity

Domain: Bias related to assessment, detection and measurement of the outcome

Appraisal: Outcome level

If the outcome is not measured in the same way in the compared groups, there is a threat to the internal validity of a study. Any differences in outcome measurements may be due to the method of measurement employed between the two groups, and not due to the

intervention/treatment of interest. Reviewers should check if the outcomes were measured in the same way. Same instrument or scale used? Same measurement timing? Same measurement procedures and instructions?

Question 9: Were outcomes measured in a reliable way?

Category: Internal validity

Domain: Bias related to assessment, detection and measurement of the outcome

Appraisal: Outcome level

Unreliability of outcome measurements is one threat that weakens the validity of inferences about the statistical relationship between the 'cause' and the 'effect' estimated in a study exploring causal effects. Unreliability of outcome measurements is one of the different plausible explanations for errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment ('cause'). Reviewers should check the details about the reliability of the measurement used, such as the number of raters, training of raters, the intra-rater and the inter-raters reliability within the study (not as reported in external sources). This question is about the reliability of the measurement performed in the study, it is not about the validity of the measurement instruments/scales used in the study. Finally, some outcomes may not rely on instruments or scales (e.g. death) and reliability of the measurements may need to be assessed in the context of the study being reviewed. [Note: Two other important threats that weaken the validity of inferences about the statistical relationship between the 'cause' and the 'effect' are low statistical power and the violation of the assumptions of statistical tests. These other two threats are explored within Question 12).]

Question 10: Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed?

Category: Internal validity

Domain: Bias related to participant retention

Appraisal: Result level

For this question, follow up refers to the period from the moment of randomization to any point in which the groups are compared during the trial. This question asks if there is complete knowledge (measurements, observations etc.) for the entire duration of the trial for all randomly allocated participants. If there is incomplete follow up from all randomly allocated participants, this is known as post-assignment attrition. As RCTs are not perfect, there is almost always post-assignment attrition, and the focus of this question is on the appropriate exploration of post-assignment attrition. If differences do exist with regards to the post-assignment attrition between the compared groups of an RCT, then there is a threat to the internal validity of that study. This is because these differences may provide a plausible alternative explanation for the observed 'effect' even in the absence of the 'cause' (the treatment or intervention of interest). It is important to note that with regards post-assignment attrition, it is not enough to know the number of participants and the proportions of participants with incomplete data; the reasons for loss to follow up are essential in the analysis of risk of bias.

Reviewers should check if there were differences with regards to the loss to follow up between the compared groups. If follow up was incomplete (incomplete information on all participants), examine the reported details about the strategies used to address incomplete follow up. This can include descriptions of loss to follow up (absolute numbers; proportions; reasons for loss to follow up) and impact analyses (the analyses of the impact of loss to follow up on results). Was there a description of the incomplete follow up including the number of participants and the specific reasons for loss to follow up? Even if follow up was incomplete, but balanced between groups, if the reasons for loss to follow up are different (e.g., side effects caused by the intervention of interest), these may impose a risk of bias if not appropriately explored in the analysis. If there are differences between groups with regards to the loss to follow up (numbers/proportions and reasons), was there an analysis of patterns of loss to follow up? If there are differences between the groups with regards to the loss to follow up, was there an analysis of the impact of the loss to follow up on the results? [Note: Question 10 is NOT about intention-to-treat (ITT) analysis; question 11 is about ITT analysis.]

Question 11: Were participants analysed in the groups to which they were randomized?

Category: Statistical conclusion validity

Appraisal: Result level

This question is about the intention-to-treat (ITT) analysis. There are different statistical analysis strategies available for the analysis of data from RCTs, such as intention-to-treat analysis (known also as intent to treat; abbreviated, ITT), per-protocol analysis, and as-treated analysis. In the ITT analysis the participants are analysed in the groups to which they were randomized. This means that regardless of whether participants received the intervention or control as assigned, were complaint with their planned assignment or participated for the entire study duration, they are still included in the analysis. The ITT analysis compares the outcomes for participants from the initial groups created by the initial random allocation of participants to those groups. Reviewers should check if an ITT analysis was reported; check the details of the ITT. Were participants analysed in the groups to which they were initially randomized, regardless of whether they participated in those groups, and regardless of whether they received the planned interventions?

[Note: The ITT analysis is a type of statistical analysis recommended in the Consolidated Standards of Reporting Trials (CONSORT) statement on best practices in trials reporting, and it is considered a marker of good methodological quality of the analysis of results of a randomized trial. The ITT is estimating the effect of offering the intervention, that is, the effect of instructing the participants to use or take the intervention; the ITT it is not estimating the effect of receiving the intervention of interest.]

Question 12: Was appropriate statistical analysis used?

Category: Statistical conclusion validity

Appraisal: Result level

Inappropriate statistical analysis may cause errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment (‘cause’). Low statistical power and the violation of the assumptions of statistical tests are two important threats that

weaken the validity of inferences about the statistical relationship between the ‘cause’ and the ‘effect’. Reviewers should check the following aspects: were the assumptions of the statistical tests were respected; if appropriate statistical power analysis was performed; if appropriate effect sizes were used; if appropriate statistical methods were used given the nature of the data and the objectives of statistical analysis (association between variables; prediction; survival analysis etc.).

Question 13: Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

Category: Statistical conclusion validity

Appraisal: Study level

The typical, parallel group RCT may not always be appropriate depending on the nature of the question being asked. Therefore, some additional RCT designs may have been employed that each come with their own additional considerations.

Crossover trials should only be conducted in people with a chronic, stable condition, where the intervention produces a short-term effect (i.e. relief in symptoms). Crossover trials should ensure there is an appropriate period of washout between treatments. This may also be considered under question 6.

Cluster RCTs randomize groups individuals or groups (e.g. communities, wards etc.) , forming ‘clusters.’ When we are assessing outcomes on an individual level in cluster trials, there are unit-of-analysis issues, as individuals within a cluster are correlated. This should be considered by the study authors when conducting analysis, and ideally authors will report the intra-cluster correlation coefficient. This may also be considered under question 12.

Stepped wedge RCTs may be appropriate to establish when and how a beneficial intervention may be best implemented within a defined setting, or due to logistical, practical, or financial considerations in the roll out of a new treatment/intervention. Data analysis in these trials should be conducted appropriately, considering the effects of time. This may also be considered under question 12.

Conclusion

Randomized controlled studies are the ideal, and often, the only included study design for systematic reviews assessing the effectiveness of interventions. All included studies must undergo rigorous critical appraisal, which in the case of quantitative study designs, is predominantly focussed on assessment of risk of bias in the conduct of the study. The revised JBI critical appraisal tool for randomized controlled trials presents an adaptable and robust new method for assessing this risk of bias. The tool has been designed to complement recent advancements in the field while maintaining its easy to follow questions. The revised JBI critical appraisal tool for randomized controlled trials offers systematic reviewers an improved and up to date method to assess the risk of bias for randomized controlled trials included in their systematic review.

References

1. Munn Z, Barker TH, Moola S, Tufanaru C, Stern C, McArthur A, et al. Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. JBI evidence synthesis. 2020; 18(10):2127-33.
2. Aromataris E, Munn Z. Chapter 1: JBI Systematic Reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. JBI; 2020.
3. Porritt K, Gomersall J, Lockwood C. JBI's systematic reviews: study selection and critical appraisal. AJN The American Journal of Nursing. 2014; 114(6):47-52.
4. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. JBI; 2020.
5. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. bmj. 2019; 366.
6. Stone JC, Glass K, Clark J, Munn Z, Tugwell P, Doi SAR. A unified framework for bias assessment in clinical research. JBI Evidence Implementation. 2019; 17(2):106-20.
7. Stone JC, Gurunathan U, Aromataris E, Glass K, Tugwell P, Munn Z, et al. Bias Assessment in Outcomes Research: The Role of Relative Versus Absolute Approaches. Value in Health. 2021; 24(8):1145-9.
8. Barker TH, Stone JC, Sears K, Klugar M, Leonardi-Bee J, Tufanaru C, et al. Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process. JBI Evidence Synthesis. 2022:10.11124.
9. Jordan Z, Lockwood C, Aromataris E, Pilla B, Porritt K, Klugar M, et al. JBI series paper 1: Introducing JBI and the JBI Model of EHBC. Journal of Clinical Epidemiology.
10. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021; 372:n160.
11. GRADE Working Group. Grading quality of evidence and strength of recommendations. Bmj. 2004; 328(7454):1490.
12. Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Completing ‘Summary of findings’ tables and grading the certainty of the evidence. Cochrane Handbook for systematic reviews of interventions. 2019:375-402.
13. Aromataris E, Stern C, Lockwood C, Barker TH, Klugar M, Jadotte Y, et al. JBI series paper 2: tailored evidence synthesis approaches are required to answer diverse questions: a pragmatic evidence synthesis toolkit from JBI. Journal of Clinical Epidemiology. 2022.
14. Stone J, Gurunathan U, Glass K, Munn Z, Tugwell P, Doi SAR. Stratification by quality induced selection bias in a meta-analysis of clinical trials. Journal of Clinical Epidemiology. 2019; 107:51-9.
15. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics. 2001; 2(4):463-71.
16. Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. J Clin Epidemiol. 2020; 117:36-45.

Appendix

Table 2 – Example presentation of results following critical appraisal using the revised JBI critical appraisal tool for randomized controlled trials.

			INTERNAL VALIDITY Bias related to:									STATISTICAL CONCLUSION VALIDITY		
		DOMAIN	Selection and Allocation			Administration of intervention/exposure			Assessment, detection, and measurement of the outcome			Participant Retention			
		QUESTION NO.	1	2	3	4	5	6	7	8	9	10	11	12	13
STUDY ID	OUTCOME	RESULT													
Study 1	Mortality	Time 1	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Mortality	Time 2										Y	Y	Y	Y
	QOL	Time 1							N	Y	Y	Y	Y	Y	Y
	QOL:	Time 2										N	Y	Y	Y
Study 2	Mortality	Time 1	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Mortality	Time 2										Y	Y	Y	Y
	QOL	Time 1							Y	Y	Y	Y	Y	Y	Y
	QOL:	Time 2										Y	Y	Y	Y

Table 2 – Example of how the results of critical appraisal may be presented when using the revised JBI critical appraisal tool for RCTs. This example has presented the results that clearly distinguishes the relationship between the result to the outcome, and the outcome to the study. Reviewers can also provide summary judgements for each domain of bias and validity construct presented. For example, for study 1, there may be a low risk of bias for the domain of 'selection and allocation' as all questions responded with a 'yes'. However, study 2, may be considered to have moderate risk of bias for the same domain, as the response to one of the questions was a 'no'.

RoB Assessor:	Date of Appraisal:	Record Number:
Study Author:	Study Title:	Study Year:

Internal Validity		Choice - Comments/Justification	Yes	No	Unclear	N/A
Bias related to selection and allocation						
1	Was true randomization used for assignment of participants to treatment groups?		☐	☐	☐	☐
2	Was allocation to treatment groups concealed?		☐	☐	☐	☐
3	Were treatment groups similar at the baseline?		☐	☐	☐	☐
Bias related to administration of intervention/exposure						
4	Were participants blind to treatment assignment?		☐	☐	☐	☐
5	Were those delivering the treatment blind to treatment assignment?		☐	☐	☐	☐
6	Were treatment groups treated identically other than the intervention of interest?		☐	☐	☐	☐
Bias related to assessment, detection and measurement of the outcome						
7	Were outcome assessors blind to treatment assignment?		Yes	No	Unclear	N/A
	Outcome 1		☐	☐	☐	☐
	Outcome 2		☐	☐	☐	☐
	Outcome 3		☐	☐	☐	☐
	Outcome 4		☐	☐	☐	☐

	Outcome 5		☐	☐	☐	☐
	Outcome 6		☐	☐	☐	☐
	Outcome 7		☐	☐	☐	☐

8	**Were outcomes measured in the same way for treatment groups?**		**Yes**	**No**	**Unclear**	**N/A**
	Outcome 1		☐	☐	☐	☐
	Outcome 2		☐	☐	☐	☐
	Outcome 3		☐	☐	☐	☐
	Outcome 4		☐	☐	☐	☐
	Outcome 5		☐	☐	☐	☐
	Outcome 6		☐	☐	☐	☐
	Outcome 7		☐	☐	☐	☐

9	**Were outcomes measured in a reliable way**		**Yes**	**No**	**Unclear**	**N/A**
	Outcome 1		☐	☐	☐	☐
	Outcome 2		☐	☐	☐	☐
	Outcome 3		☐	☐	☐	☐
	Outcome 4		☐	☐	☐	☐
	Outcome 5		☐	☐	☐	☐
	Outcome 6		☐	☐	☐	☐

	Outcome 7		☐	☐	☐	☐

Bias related to participant retention

10	**Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed?**					
	Outcome 1		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 2		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 3		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 4		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐

Outcome 5		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 6		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 7		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐

Statistical Conclusion Validity

11	**Were participants analysed in the groups to which they were randomized?**					
Outcome 1		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 2		**Yes**	**No**	**Unclear**	**N/A**

Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 3		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 4		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 5		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 6		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 7		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐

	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐

12	**Was appropriate statistical analysis used?**					
	Outcome 1		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 2		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 3		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 4		**Yes**	**No**	**Unclear**	**N/A**
	Result 1		☐	☐	☐	☐
	Result 2		☐	☐	☐	☐
	Result 3		☐	☐	☐	☐
	Outcome 5		**Yes**	**No**	**Unclear**	**N/A**

Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 6		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐
Outcome 7		**Yes**	**No**	**Unclear**	**N/A**
Result 1		☐	☐	☐	☐
Result 2		☐	☐	☐	☐
Result 3		☐	☐	☐	☐

		Yes	**No**	**Unclear**	**N/A**
13	**Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?**	☐	☐	☐	☐

Overall appraisal: **Include:** ☐ **Exclude:** ☐ **Seek Further Info:** ☐

Comments:

Table 3 – The JBI Critical Appraisal Tool for RCTs