RoB 2: a revised tool for assessing risk of bias in randomised trials

Journal:	*BMJ*
Manuscript ID	BMJ-2019-049848.R1
Article Type:	Research methods and reporting
BMJ Journal:	BMJ
Date Submitted by the Author:	12-Jun-2019
Complete List of Authors:	Sterne, Jonathan; University of Bristol, Population Health Sciences; University of Bristol Savovic , Jelena; University of Bristol, School of Social & Community Medicine Page, Matthew; Monash University, School of Public Health and Preventive Medicine Elbers, Roy; University of Bristol, Population Health Sciences Blencowe, Natalie; University of Bristol, Centre for Surgical Research, School of Social and Community Medicine Boutron, Isabelle ; Center of Research in Epidemilogy and Statistics Sorbone Paris, CRESS UMR 1153, METHODS team, Cates, Chris; University of London Saint George's Cheng, Vincent; University of Bristol, Population Health Sciences Corbett, Mark; University of York, Centre for Reviews and Dissemination Eldridge, Sandra; Queen Mary, University of London, Primary Care and Public Health Emberson, Jonathan; University of Oxford, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health Hernan, Miguel; Harvard T.H.Chan School of Public Health, Hopewell, Sally; Centre for Statistics in Medicine, Oxford Clinical Trials Research Unit Hrobjartsson, Asbjørn; The Nordic Cochrane Centre, Junqueira, Daniela; University of Alberta, Department of Emergency Medicine, Faculty of Medicine and Dentistry Juni, Peter; University of Toronto, St. Michael's Hospital, Applied Health Research Centre (AHRC) Kirkham, Jamie J.; University of Liverpool, Biostatistics Lasserson, Toby; Cochrane, Cochrane Editorial Unit Li, T; Johns Hopkins University, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health McAleenan, Alexandra; University of Bristol, Population Health Sciences Reeves, Barnaby; University of Bristol, Bristol Heart Institute Shepperd, Sasha; University of Oxford, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health Shrier, Ian; SMBDJewish General Hospital Does not like open peer review, Centre for Clinical Epidemiology and Communit Stewart, Lesley; University of York, Centre for Reviews and

	Dissemination Tilling, Kate; University of Bristol, Population Health Sciences, Bristol Medical School White, Ian; MRC Biostatistics Unit, Whiting, Penny; NIHR CLAHRC West, University Hospitals Bristol NHS Foundation Trust; University of Bristol, School of Social and Community Medicine Higgins, Julian; University of Bristol, School of Social and Community Medicine
Keywords:	systematic reviews, meta-analysis, risk of bias, randomized trials, intention to treat, per-protocol

RoB 2: a revised tool for assessing risk of bias in randomised trials

Jonathan AC Sterne, professor, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK

Jelena Savović, senior research fellow, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR CLAHRC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

Matthew J Page, research fellow, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Roy G Elbers, senior research associate, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

Natalie S Blencowe, National Institute for Health Research clinical lecturer, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK

Isabelle Boutron, professor, METHODS team, Epidemiology and Biostatistics Center, INSERM UMR 1153, Paris, France; Paris Descartes University, Paris, France; and Cochrane France, Paris, France

Christopher J Cates, senior clinical research fellow, Population Health Research Institute, St George's, University of London, London, UK

Hung-Yuan Cheng, senior research associate, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK

Mark S Corbett, research fellow, Centre for Reviews and Dissemination, University of York, York, UK

Sandra M Eldridge, professor, Pragmatic Clinical Trials Unit, Centre for Primary Care and Public Health, Queen Marys University of London, UK

Jonathan R Emberson, associate professor, Medical Research Council Population Heath Research Unit, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Miguel A Hernán, professor, Departments of Epidemiology and Biostatistics, Harvard T H Chan School of Public Health, Harvard-MIT Division of Health Sciences of Technology, Boston, Massachusetts, USA

Sally Hopewell, associate professor, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Asbjørn Hróbjartsson, professor, Centre for Evidence-Based Medicine Odense (CEBMO), Odense University Hospital, Odense, Denmark; and Department of Clinical Research, University of Southern Denmark, Odense, Denmark; and Open Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark

Daniela R Junqueira, research associate, Department of Emergency Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada

Peter Jüni, professor, Applied Health Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Department of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

Jamie J Kirkham, professor, Centre for Biostatistics, University of Manchester, Manchester, UK

Toby Lasserson, senior editor, Editorial & Methods Department, Cochrane Central Executive, London, UK
Tianjing Li, associate professor, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
Alexandra McAleenan, senior research associate, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
Barnaby C Reeves, professorial research fellow, Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK
Sasha Shepperd, professor, Nuffield Department of Population Health, University of Oxford, Oxford, UK
Ian Shrier, investigator, Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada
Lesley A Stewart, professor, Centre for Reviews and Dissemination, University of York, York, UK
Kate Tilling, professor, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK
Ian R White, professor, MRC Clinical Trials Unit at University College London, London, UK
Penny F Whiting, senior lecturer, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; and NIHR CLAHRC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK
Julian PT Higgins, professor, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; NIHR CLAHRC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK; and NIHR Bristol Biomedical Research Centre, Bristol, UK

Correspondence to: Prof Jonathan AC Sterne, Population Health Sciences, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. Email: jonathan.sterne@bristol.ac.uk.

Key messages

- Assessment of risk of bias is regarded as an essential component of a systematic review on the effects of an intervention. The most commonly used tool for assessing risk of bias in randomised trials is the Cochrane Risk of Bias tool, which was introduced in 2008.
- Potential improvements to the Cochrane risk of bias tool were identified based on reviews of the literature, user experience and feedback, approaches used in other risk of bias tools, and recent developments in estimation of intervention effects from randomised trials.
- We developed and piloted a revised tool for assessing risk of bias in randomised trials (RoB 2).
- Bias is addressed in five distinct domains, within each of which answers to signalling questions lead to judgements of "low risk of bias", "some concerns" or "high risk of bias".
- The judgements within each domain lead to an overall risk of bias judgement for the result being assessed. This should facilitate stratification of meta-analyses according to risk of bias.

Standfirst

Assessment of risk of bias is regarded as an essential component of a systematic review on the effects of an intervention. The most commonly used tool for randomised trials is the Cochrane Risk of Bias tool. We updated the tool to respond to developments in understanding how bias arises in randomized trials, and to address user feedback on and limitations of the original tool.

Introduction

An evaluation of the risk of bias in each study included in a systematic review documents potential flaws in the evidence summarised and contributes to the certainty in the overall evidence.[1] The Cochrane tool for assessing risk of bias in randomised trials ('RoB tool')[2] has been widely used in both Cochrane and other systematic reviews, with over 40,000 citations in Google Scholar.

Many innovative characteristics of the original RoB tool have been widely accepted. It replaced the notion of assessing study quality with that of assessing risk of bias (we define bias as a systematic deviation from the effect of intervention that would be observed in a large randomized trial without any flaws). Quality is not well defined and can include study characteristics (such as performing a sample size calculation) that are not inherently related to bias in the study's results. The RoB tool considers biases arising at different stages of a trial ('bias domains'), which were chosen based on both empirical evidence and theoretical considerations. Assessments of risk of bias are supported by quotes from sources describing the trial (e.g. trial protocol, registration record, results report) or by justifications written by the assessor.

After nearly a decade of experience of using the RoB tool, potential improvements have been identified. A formal evaluation found some bias domains to be confusing at times, with assessment of bias due to incomplete outcome data and selective reporting of outcomes causing particular difficulties, and confusion over whether studies that were not blinded should automatically be considered to be at high risk of bias.[3] More guidance on incorporating RoB assessments into meta-analyses and review conclusions is also needed.[4 5] A review of comments and user practice found that both Cochrane and non-Cochrane systematic reviews often implemented the RoB tool in non-standard ways.[6] Few trials are assessed as at low risk of bias, and judgements of 'unclear' risk of bias are common.[6 7] Empirical studies have found only moderate reliability of risk-of-bias judgements.[8]

We developed a revised risk-of-bias assessment tool that addresses these issues, as well as incorporating advances in assessment of risk of bias used in other recently developed tools[9 10] and integrating recent developments in estimation of intervention effects from randomised trials.[11]

Development of the revised RoB tool

We followed the principles adopted for development of the original RoB tool and for the ROBINS-I tool for assessing risk of bias in non-randomised studies of interventions.[2 9] A core

group coordinated development of the tool, including recruitment of collaborators, preparation and revision of documents, and administrative support.

Preliminary work included a review of how the original tool was used in practice,[6] a systematic review and meta-analysis of meta-epidemiological studies of empirical evidence for biases associated with characteristics of randomised trials[12] and a cross-sectional study of how selective outcome reporting was assessed in Cochrane reviews.[13] We also drew on a systematic review of the theoretical and conceptual literature on types of bias in epidemiology, which sought papers and textbooks presenting classifications or definitions of biases, and organized these into a coherent framework (paper in preparation).

The core group developed an initial proposal and presented it, together with the latest empirical evidence of biases in randomised trials, at a meeting in August 2015 attended by 24 contributors. Meeting participants agreed on the methodological principles underpinning the new tool and the bias domains to be addressed, and formed working groups for each domain. The groups were tasked with developing 'signalling questions' (reasonably factual questions with yes/no answers that inform risk-of-bias judgements), together with guidance for answering these questions and broad considerations for how to judge the risk of bias for the domain.

The materials prepared by the working groups were assembled and edited by the core team and the resulting draft was piloted by experienced and novice systematic reviewers during a 3-day event in February 2016, with 17 participants present and 10 participants contributing remotely. Issues identified in the pilot were documented and addressed in a new draft discussed at a second development meeting in April 2016, also attended by 24 contributors. Subsequently, working groups developed criteria for reaching domain-level risk-of-bias judgements based on answers to signalling questions, and expanded the guidance. The core team designed algorithms to match the criteria, which were checked by the working groups. The resulting revision was tested in another round of piloting by 10 systematic review authors during mid-2016.

A complete draft of RoB version 2 ('RoB 2'), together with detailed guidance, was posted at www.riskofbias.info in October 2016, coinciding with the Cochrane Colloquium in Seoul, South Korea. Feedback was invited through direct contact with the development group. Several review teams subsequently piloted the draft tool and provided feedback. Further modifications, particularly improvements in wording and clarity, splitting compound signalling questions, adding new questions and addressing methodological issues, were made based on feedback from training events (including webinars) conducted between 2016 and 2018, as well as individual feedback from users across the world.

Version 2 of the Cochrane tool for assessing risk of bias in randomised trials

RoB 2 provides a framework for assessing the risk of bias in a single estimate of an intervention effect reported from a randomised trial. The effect assessed is a comparison of two interventions, which we refer to as the experimental and comparator interventions, for a specific outcome or endpoint. The process of making a RoB 2 assessment is summarised in Figure 1. Preliminary considerations (Box 1) include specifying which result is being assessed, specifying how this result is being interpreted (see 'The intervention effect of interest' below) and listing the sources of information used to inform the assessment. Review authors should contact trial authors should obtain information that is omitted from published and online sources, so far as this is feasible. Note that RoB assessments may be needed for results relating to multiple outcomes from the included trials.

RoB 2 is structured into five bias domains, listed in Table 1. The domains were selected to address all important mechanisms by which bias can be introduced into the results of a trial, based on a combination of empirical evidence and theoretical considerations. We did not include domains for features that would be expected to operate indirectly, through the included bias domains.[14 15] For this reason, we excluded some trial features, such as funding source and single- versus multi-centre status, that have been associated empirically with trial effect estimates from trials.

We label the domains using descriptions of the causes of bias addressed, avoiding terms used in the original RoB tool (such as 'selection bias' and 'performance bias') because they are used inconsistently or not known by many people outside Cochrane.[16] Each domain is mandatory, and no others can be added, although we have developed versions of RoB 2 that deal with additional issues that arise in trials with cluster-randomised or crossover designs (see www.riskofbias.info). Within each domain, the assessment comprises:

1. a series of 'signalling questions';
2. a judgement about risk of bias for the domain, facilitated by an algorithm that maps responses to signalling questions to a proposed judgement;
3. free text boxes to justify responses to the signalling questions and risk-of-bias judgements; and
4. (optional) free text boxes to predict (and explain) the likely direction of bias.

Table 2 lists the most important changes made in RoB 2, compared with the original Cochrane RoB tool.

Signalling questions

Signalling questions aim to elicit information relevant to an assessment of risk of bias and are shown in Table 1. The questions seek to be reasonably factual in nature. The response options are 'Yes', 'Probably yes', 'Probably no', 'No' and 'No information'. To maximise their simplicity and clarity, signalling questions are phrased such that 'Yes' may indicate either lower or higher risk of bias, depending on the most natural way to ask the question. The online supplementary material includes elaborations providing guidance on how to answer each question.

Responses of 'Yes' and 'Probably yes' have the same implications for risk of bias, as do responses of 'No' and 'Probably no'. 'Yes' and 'No' typically imply that firm evidence is available; the 'Probably' responses typically imply that a judgement has been made. Where there is a need to distinguish between "Some concerns" and "High risk of bias" this is dealt with through an additional signalling question, rather than by making a distinction between responses "Probably yes" and "Yes", or between "Probably no" and "No". The 'No

information' response should be used only when insufficient details are available to permit a different response, and when in the absence of these details it would be unreasonable to respond 'Probably yes' or 'Probably no'. For example, in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomisation sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information' to the signalling question about sequence generation (the rationale for the response should be provided in the free text box). Some signalling questions are answered only if the response to a previous question indicates they are required.

The intervention effect of interest

Assessments for the domain 'Bias due to deviations from intended interventions' differ according to whether review authors are interested in quantifying: (1) the effect of assignment to the interventions at baseline regardless of whether the interventions are received during follow-up (the 'intention-to-treat effect'); or (2) the effect of adhering to the interventions as specified in the trial protocol (the 'per-protocol effect'). These effects will differ if some patients do not receive their assigned intervention or deviate from the assigned intervention after baseline. Each may be of interest.[11] For example, the effect of assignment to intervention may be appropriate to inform a health policy question about whether to recommend an intervention (e.g. a screening programme) in a particular health system, whereas the effect of adhering to the intervention more directly informs a care decision by an individual patient (e.g. whether to be screened). Changes to an intervention that are consistent with the trial protocol (even if not explicitly discussed in the protocol), such as cessation of a drug because of toxicity or switch to second-line chemotherapy because of progression of cancer, do not cause bias and should not be considered to be deviations from intended intervention.

The effect of assignment to intervention should be estimated by an intention-to-treat (ITT) analysis that includes all randomised participants.[17] However, estimates of per-protocol effects commonly used in reports of randomised trials are problematic and may be seriously biased.[18] These include estimates from naïve 'per protocol' analyses restricted to individuals

who adhered to their assigned intervention, and 'as-treated' analyses in which participants are analysed according to the intervention they received, even if their assigned group is different. These approaches are problematic because prognostic factors may influence whether individuals receive their allocated intervention. It is possible to use data from a randomised trial to derive an unbiased estimate the effect of adhering to intervention.[19 20] However, appropriate methods require strong assumptions and published applications are relatively rare to date. For trials comparing interventions that are sustained over time, appropriate methods also require measurement of and adjustment for both the pre- and post-randomisation prognostic factors that predict deviations from intervention.[11] For these reasons, most systematic reviews are likely to address the effect of assignment rather than adherence to intervention.

Risk-of-bias judgements

The risk-of-bias judgements for each domain are 'Low risk of bias', 'Some concerns' or 'High risk of bias'. Judgements are based on, and summarise, the answers to signalling questions. Review authors should interpret 'risk of bias' as 'risk of material bias': concerns should be expressed only about issues likely to have a notable effect on the result being assessed.

An important innovation in RoB 2 is the inclusion of algorithms that map responses to signalling questions to a proposed risk-of-bias judgement for each domain (see online supplementary material). Review authors can override these proposed judgements if they feel it is appropriate to do so.

Free text boxes alongside the signalling questions and judgements allow assessors to provide support for the responses. Brief direct quotations from the texts of the study reports (including trial protocols) should be used whenever possible, supplemented by any information obtained from authors when contacted. Reasons for any judgements that do not follow the algorithms should be provided. RoB 2 includes optional judgements of the direction of the bias for each domain and overall. If review authors do not have a clear rationale for judging the likely direction of the bias, they should not guess it.

Overall risk of bias for the result

The response options for an overall risk-of-bias judgement are the same as for individual domains. Table 3 shows the approach to mapping bias judgements within domains to an overall judgement for the result. The overall risk of bias generally corresponds to the worst risk of bias in any of the domains. However, if a study is judged to have 'some concerns' about risk of bias for multiple domains, it may be judged as at high risk of bias overall. Figure 2 shows a forest plot that displays domain-specific and overall risk of bias, with the meta-analysis stratified by overall risk of bias.

Discussion

We have substantially revised the Cochrane tool for assessing risk of bias in the results of randomised trials, in order to address limitations identified since it was published in 2008 and incorporate improvements that aim to increase the reliability of assessments. RoB 2 is based on wide consultation within and outside Cochrane, extensive piloting and integration of feedback based on user experience. Assessments are made within five bias domains, within which answers to signalling questions address a broader range of issues than in the original RoB tool. These include whether post-randomization deviations from intervention caused bias in trials in which blinding was either not feasible or not implemented and whether outcome data were missing for reasons likely to lead to bias. Assessment of selective reporting is focussed on a reported result for an outcome, rather than selective non-reporting of other outcomes that were measured in the trial. RoB 2 also incorporates recent developments in estimation of intervention effects from randomised trials: we distinguish bias in the effect of assignment to interventions from bias in the effect of adhering to the interventions as specified in the trial protocol.[11]

RoB 2 addresses the risk of bias in a single estimate of intervention effect for a single outcome or endpoint, rather than for a whole trial. This is because risk of bias is outcome-specific for domains such as bias in measurement of the outcome, and may be specific to a particular estimate (e.g. when both intention-to-treat and 'per protocol' analyses have been conducted). We recommend that overall RoB 2 judgements of risk of bias for individual results

should be the primary means of distinguishing stronger from weaker evidence in the context of a meta-analysis (or other synthesis) of randomised trials. They should also influence the strength of conclusions drawn from a systematic review (potentially as part of a GRADE assessment)[21]. We strongly encourage stratification by overall risk-of-bias judgement as a default meta-analysis strategy, as shown in Figure 2. To facilitate this, we suggest that systematic review preparation software provides data fields for risk-of-bias assessments. We are preparing an interactive web tool for completing RoB 2 assessments, which we hope will interface well with other systematic review software.

In RoB 2, judgements about risk of bias are derived by algorithms based on answers to specific 'signalling' questions. The added structure provided by the signalling questions aims to make the assessment easier and more efficient to use, as well as to improve agreement between assessors. We believe this approach to be more straightforward than the direct judgements about risk of bias required in the original RoB tool. The algorithms include explicit mappings for situations in which there is no information to answer a signalling question, which do not necessarily map to a negative assessment of the trial. For example, when randomisation methods are described and are adequate, the response to the signalling question about baseline imbalances between intervention groups leads to low risk of bias either when such imbalances are compatible with chance, or when there is no information about baseline imbalances. We removed the option of an 'Unclear' judgement in favour of a graded set of response options (from 'Low' to 'Some concerns' to 'High'). We envisage that systematic reviews will report the domain-level and overall risk-of-bias judgements in tables or figures contained in the main review text. In addition, we encourage reporting of answers to signalling questions, together with direct quotes from papers and free-text justification of the answers, in an appendix.

We expect the refinements we have made to the RoB tool to lead to a greater proportion of trial results being assessed as at low risk of bias, because our algorithms map some circumstances to 'Low' risk of bias when users of the previous tool would typically have assessed them to be at 'Unclear' (or even 'High') risk of bias. This is particularly the case for trials that are not blinded, where risk of bias in the effect of assignment to intervention may

be low despite many users of the original RoB tool assigning 'High' risk of bias to this domain. We believe that judgements of low risk of bias should be readily achievable for a randomised trial, a study design that is scientifically strong, well understood and often well implemented in practice. We hope that Version 2 of the Cochrane Risk-of-Bias tool (RoB 2) will be useful to systematic review authors and those making use of reviews, by providing a coherent framework for understanding and identifying trials at risk of bias. This framework may also help those designing, conducting and reporting randomised trials to achieve the most reliable findings possible.

Dedication

We dedicate this work to Professor Douglas G Altman, whose contributions were of fundamental importance to development of risk-of-bias assessment in systematic reviews.

Acknowledgements

We thank Henning Keinke Andersen, Nancy Berkman, Mike Campbell, Rachel Churchill, Mike Clarke, Nicky Cullen, Francois Curtin, Amy Drahota, Bruno Giraudeau, Jeremy Grimshaw, Sharea Ijaz, Yoon Loke, Geraldine Macdonald, Richard Morris, Mona Nasser, Nishith Patel, Jani Ruotsalainen, Stephen Senn, Holger Schünemann, Nandi Siegfried, Jayne Tierney and Sunita Vohra for contributing to discussions, and we thank Andrew Beswick, Julia Bidonde, Angela Busch, staff at Cochrane Argentina, Karen Dawe, Franco De Crescenzo, Kristine Egberts, Clovis Mariano Faggion Jr, Clare French, Lina Gölz, Valerie Hoffman, Joni Jackson, Tim Jones, Kayleigh Kew, Elsa Marques, Silvia Minozzi, Theresa Moore, Rebecca Normansell, Rosanne Freak-Poli, Sarah Lensen, José López-López, Marlies Manders, Luke McGuinness, Spyros Papageorgiou, Melissa Randall, Phil Riley, Claudia Smeets, Meera Viswanathan and Tanya Walsh for contributing to piloting of earlier drafts of the RoB 2 tool.

Funding

The development of the RoB 2 tool was supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- N61), with the support of the host MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures - MR/K025643/1), by MRC research grant MR/M025209/1, and by a grant from The Cochrane Collaboration. Sterne, Eldridge and Higgins are National Institute for Health Research (NIHR) Senior Investigators. Sterne, Blencowe, Cheng, Reeves and Higgins are supported by NIHR Bristol Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. Sterne and Higgins are members of the MRC Integrative Epidemiology Unit at the University of Bristol. Savović, Whiting and Higgins are supported by the NIHR Collaboration for Leadership in Applied Health Research and Care West (CLAHRC West) at University Hospitals Bristol NHS Foundation Trust. Jüni is a Tier 1 Canada Research Chair in Clinical Epidemiology of Chronic Diseases supported by the Canada Research Chairs Programme. Page is supported by an Early Career Fellowship from the Australian National Health and Medical Research Council (NHMRC 1088535). White was supported by the Medical Research Council Programme MC_UU_12023/21. The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR, MRC, the NHMRC or the Department of Health and Social Care.

Contributions of authors

JACS, JS and JPTH conceived the project. JACS, JPTH, JS, MJP and RGE oversaw the project. JACS, JS, AH, IB, BCR and JJK led working groups. All authors contributed to development of RoB 2 and to writing associated guidance. JACS, JS and JPTH wrote the first draft of the manuscript. All authors reviewed and commented on drafts of the manuscript.

Provenance

The authors are epidemiologists, statisticians, systematic reviewers, trialists and health services researchers, many of whom are involved with Cochrane systematic reviews, methods groups and training events. Development of RoB 2 was informed by relevant methodological

literature, previously published tools for assessing methodological quality of randomized trials, systematic reviews of such tools and relevant literature, and by the authors' experience of developing tools to assess risk of bias in randomized and non-randomized studies, diagnostic test accuracy studies and systematic reviews. All authors contributed to development of RoB 2 and to writing associated guidance. All authors reviewed and commented on drafts of the manuscript. Jonathan Sterne will act as guarantor.

Patient and Public Involvement

Patients and the public were not involved in this methodological research. We plan to disseminate the research widely, including to community participants in Cochrane.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and we expect them to declare no support from any organisation for the submitted work; 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.

References

1. 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(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026 [published Online First: 2011/01/05]
2. Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
3. Savović J, Weeks L, Sterne JAC, 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
4. Hopewell S, Boutron I, Altman DG, et al. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *BMJ Open* 2013;3(8):e003342. doi: 10.1136/bmjopen-2013-003342 [published Online First: 2013/08/27]
5. Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *J Epidemiol Community Health* 2015;69(2):189-95. doi: 10.1136/jech-2014-204711 [published Online First: 2014/12/08]
6. Jorgensen L, Paludan-Muller 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(1):80. doi: 10.1186/s13643-016-0259-8
7. Dechartres A, Trinquart L, Atal I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ* 2017;357:j2490. doi: 10.1136/bmj.j2490 [published Online First: 2017/06/10]
8. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *Journal of Clinical Epidemiology* 2013;66(9):973-81. doi: 10.1016/j.jclinepi.2012.07.005 [published Online First: 2012/09/18]
9. Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ* 2016;355:i4919.
10. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009
11. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med* 2017;377(14):1391-98. doi: 10.1056/NEJMsm1605385 [published Online First: 2017/10/05]
12. Page MJ, Higgins JPT, Clayton G, et al. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. *PLoS One* 2016;11(7):e0159267. doi: 10.1371/journal.pone.0159267 [published Online First: 2016/07/12]
13. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Syst Rev* 2016;5(1):108. doi: 10.1186/s13643-016-0289-2 [published Online First: 2016/07/09]

14. Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. *The Cochrane database of systematic reviews* 2017;2:MR000033. doi: 10.1002/14651858.MR000033.pub3 [published Online First: 2017/02/17]
15. Bafeta A, Dechartres A, Trinquart L, et al. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ* 2012;344:e813. doi: 10.1136/bmj.e813 [published Online First: 2012/02/16]
16. Mansournia MA, Higgins JPT, Sterne JAC, et al. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology* 2016 (Published online 29 September) doi: 10.1097/EDE.0000000000000564
17. Fergusson D, Aaron SD, Guyatt G, et al. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325(7365):652-4. doi: 10.1136/bmj.325.7365.652 [published Online First: 2002/09/21]
18. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9(1):48-55. doi: 10.1177/1740774511420743 [published Online First: 2011/09/29]
19. Murray EJ, Hernan MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials* 2016;13(4):372-8. doi: 10.1177/1740774516634335 [published Online First: 2016/03/10]
20. Lodi S, Sharma S, Lundgren JD, et al. The per-protocol effect of immediate versus deferred antiretroviral therapy initiation. *AIDS* 2016;30(17):2659-63. doi: 10.1097/QAD.0000000000001243 [published Online First: 2016/10/27]
21. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-26.

Box 1. The RoB 2 tool: Preliminary considerations

For the purposes of this assessment, the interventions being compared are defined as

Experimental: | Comparator:

Specify which outcome is being assessed for risk of bias

Specify the numerical result being assessed. In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Is the review team's aim for this result…?

☐ to assess the effect of *assignment to intervention* (the 'intention-to-treat' effect)
☐ to assess the effect of *adhering to intervention* (the 'per-protocol' effect)

If the aim is to assess the effect of *adhering to intervention*, select the deviations from intended intervention that should be addressed (at least one must be checked):

☐ occurrence of non-protocol interventions
☐ failures in implementing the intervention that could have affected the outcome
☐ non-adherence to their assigned intervention by trial participants

Which of the following sources were <u>obtained</u> to help inform the risk-of-bias assessment? (tick as many as apply)

☐ Journal article(s)
☐ Trial protocol
☐ Statistical analysis plan (SAP)
☐ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
☐ Company-owned trial registry record (e.g. GSK Clinical Study Register record)
☐ "Grey literature" (e.g. unpublished thesis)
☐ Conference abstract(s) about the trial
☐ Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
☐ Research ethics application
☐ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)
☐ Personal communication with trialist
☐ Personal communication with the sponsor

Table 1. Version 2 of the Cochrane Risk-of-bias assessment tool for randomised trials: bias domains, signalling questions, response options and risk of bias judgements. Signalling questions for bias due to deviations from intended interventions relate to the effect of assignment to intervention.

Bias arising from the randomisation process	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 randomisation 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?	
Bias due to deviations from intended interventions	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 experimental context?	NA/Y/PY/PN/N/NI
	2.4 If Y/PY/NI to 2.3: Were these deviations likely to have affected the outcome?	NA/Y/PY/PN/N/NI
	2.5. If Y/PY 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 randomised?	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?	

Bias due to missing outcome data	3.1 Were data for this outcome available for all, or nearly all, participants randomised?	Y/PY/PN/N/NI
	3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	NA/Y/PY/PN/N
	3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	NA/Y/PY/PN/N/NI
	3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	NA/Y/PY/PN/N/NI
	Risk of bias judgement (Low/High/Some concerns)	
	Optional: What is the predicted direction of bias due to missing outcome data?	
Bias in measurement of the outcome	4.1 Was the method of measuring the outcome inappropriate?	Y/PY/PN/N/NI
	4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Y/PY/PN/N/NI
	4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Y/PY/PN/N/NI
	4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	NA/Y/PY/PN/N/NI
	4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	NA/Y/PY/PN/N/NI
	Risk of bias judgement (Low/High/Some concerns)	
	Optional: What is the predicted direction of bias in measurement of the outcome?	
Bias in selection of the reported result	5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalised before unblinded outcome data were available for analysis? Is the numerical result being assessed likely to have been selected, on the basis of the results, from...	Y/PY/PN/N/NI
	5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	Y/PY/PN/N/NI
	5.3 ... multiple eligible analyses of the data?	Y/PY/PN/N/NI
	Risk of bias judgement (Low/High/Some concerns)	
	Optional: What is the predicted direction bias due to selection of the reported results?	
Overall bias	**Risk of bias judgement** (Low/High/Some concerns)	
	Optional: What is the overall predicted direction of bias for this outcome?	

Table 2. Major changes in version 2 of the Cochrane Risk-of-bias assessment tool, compared with the original Cochrane Risk-of-bias tool

Bias domain	Major changes compared with the original RoB tool
1. Bias arising from the randomisation process	The original tool did not address issues relating to baseline differences. We emphasise that baseline differences that are compatible with chance do not lead to a risk of bias
2. Bias due to deviations from intended interventions	1. The original tool only addressed whether participants, carers and people delivering the interventions were aware of participants' assigned intervention during the trial. The revised tool recognises that open trials can be at low risk of bias, if there were no deviations from intended intervention that arose because of the experimental context. 2. Whether the analysis was appropriate to estimate the effect of assignment to intervention was previously assessed in relation to missing outcome data. 3. The original tool did not address bias in estimating the effect of adhering to intervention. Imbalances in co-interventions, failures in implementing the intervention and non-adherences can all bias such estimates. An appropriate analysis has the potential to address such biases, in some circumstances.
3. Bias due to missing outcome data	1. Issues relating to exclusions in analyses (for example naïve 'per-protocol' analyses) are now addressed in the 'deviations from intended intervention' domain. 2. Whether missing outcome data lead to bias depends on the relationship between the true value of the outcome in participants with missing outcome data, and the 'missingness mechanism' (the process that led to outcome data being missing). This domain has been substantially reworked, to reflect situations in which missing outcome data do and do not lead to bias in a 'complete case' analysis. 3. We clarify that multiple imputation methods will not remove or reduce bias that occurs when missingness in the outcome depends on its true value, unless such missingness can be explained by measured variables.
4. Bias in measurement of the outcome	The original tool only addressed whether outcome assessors were aware of the intervention received by study participants. This domain now covers a range of ways in which the method of outcome measurement can lead to bias, including issues related to passive detection of outcomes that may be particularly relevant for adverse effects (harms) of interventions.

5. Bias in selection of the reported result	1. Unlike the original tool, this domain does not address bias due to selective *non-reporting* of results (either because of non-publication of whole studies or selective reporting of outcomes) for outcome domains that were measured and analysed. Such bias puts the result of a *synthesis* at risk because results are omitted based on their direction, magnitude or statistical significance. It should therefore be addressed at the review level, as part of an integrated assessment of the risk of reporting bias. 2. A judgment of low risk of bias requires that the trial was analysed in accordance with a pre-specified plan that was finalised before unblinded outcome data were available for analysis.

Table 3. Reaching an overall risk-of-bias judgement for a specific result.

Overall risk-of-bias judgement	Criteria
Low risk of bias	The study is judged to be at **low risk of bias for all domains** for this result.
Some concerns	The study is judged to raise **some concerns** in at least one domain for this result, but not to be at high risk of bias for any domain.
High risk of bias	The study is judged to be at **high risk of bias** in at least one domain for this result. Or The study is judged to have **some concerns** for **multiple domains** in a way that substantially lowers confidence in the result.

Figure 1: Summary of the process of assessing risk of bias in a systematic review of randomized trials

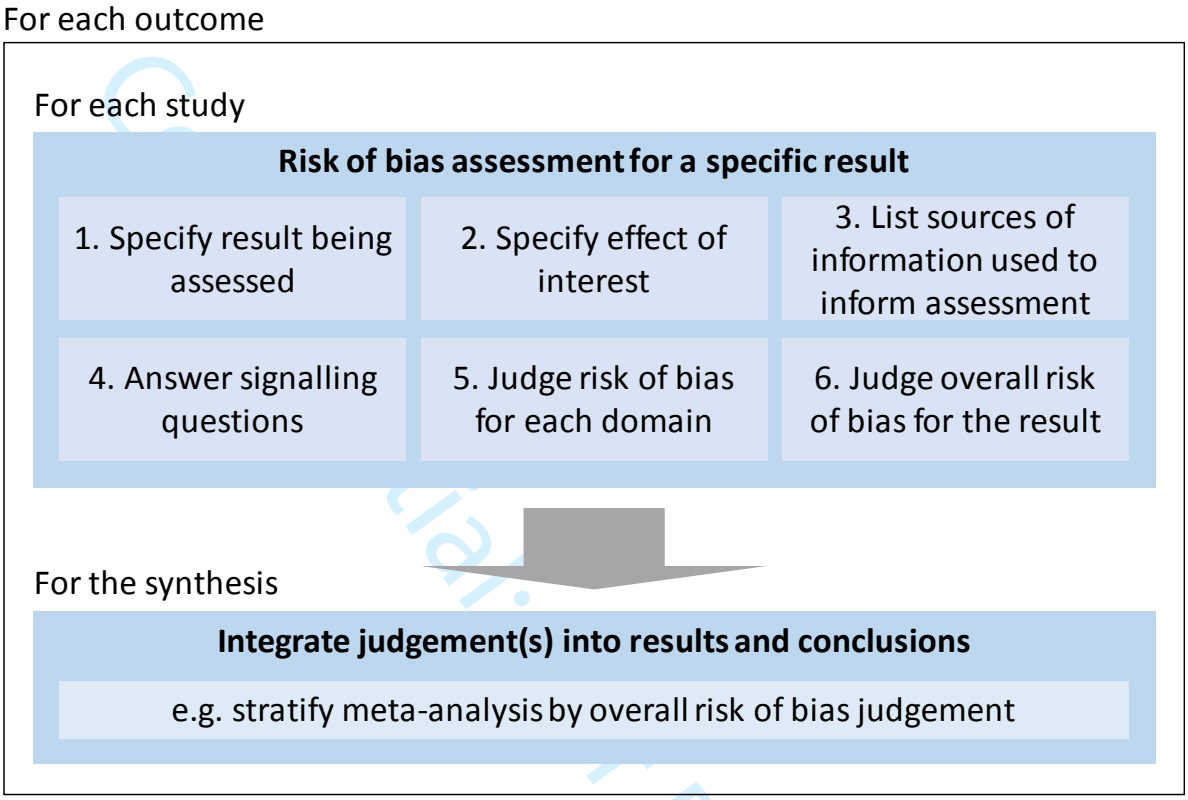
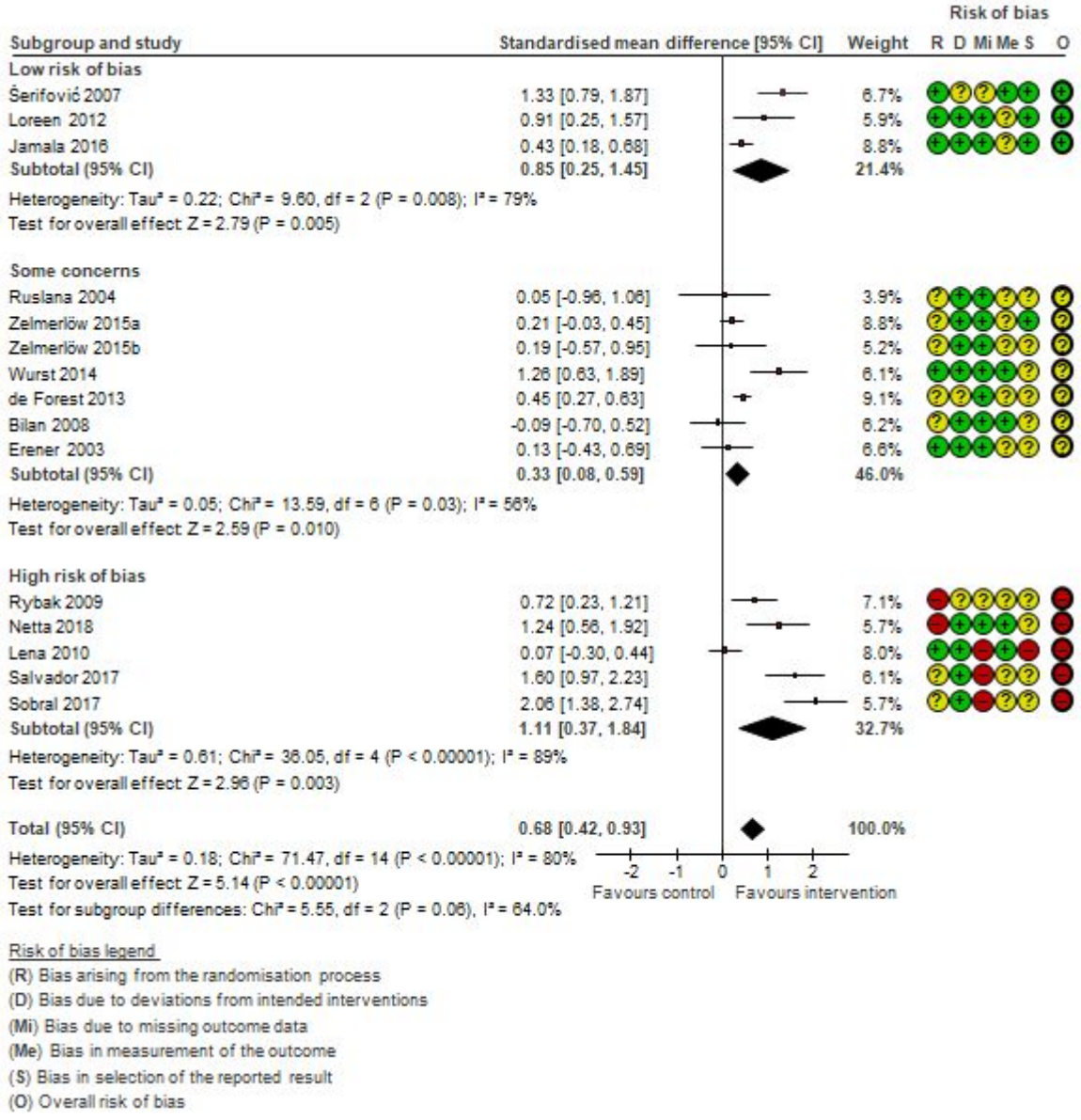


Figure 2: Example of a forest plot showing results of a RoB 2 assessment. Studies are stratified by overall risk of bias.



Revised Cochrane risk-of-bias tool for randomized trials (RoB 2): supplementary material

Domain 1: Risk of bias arising from the randomization process

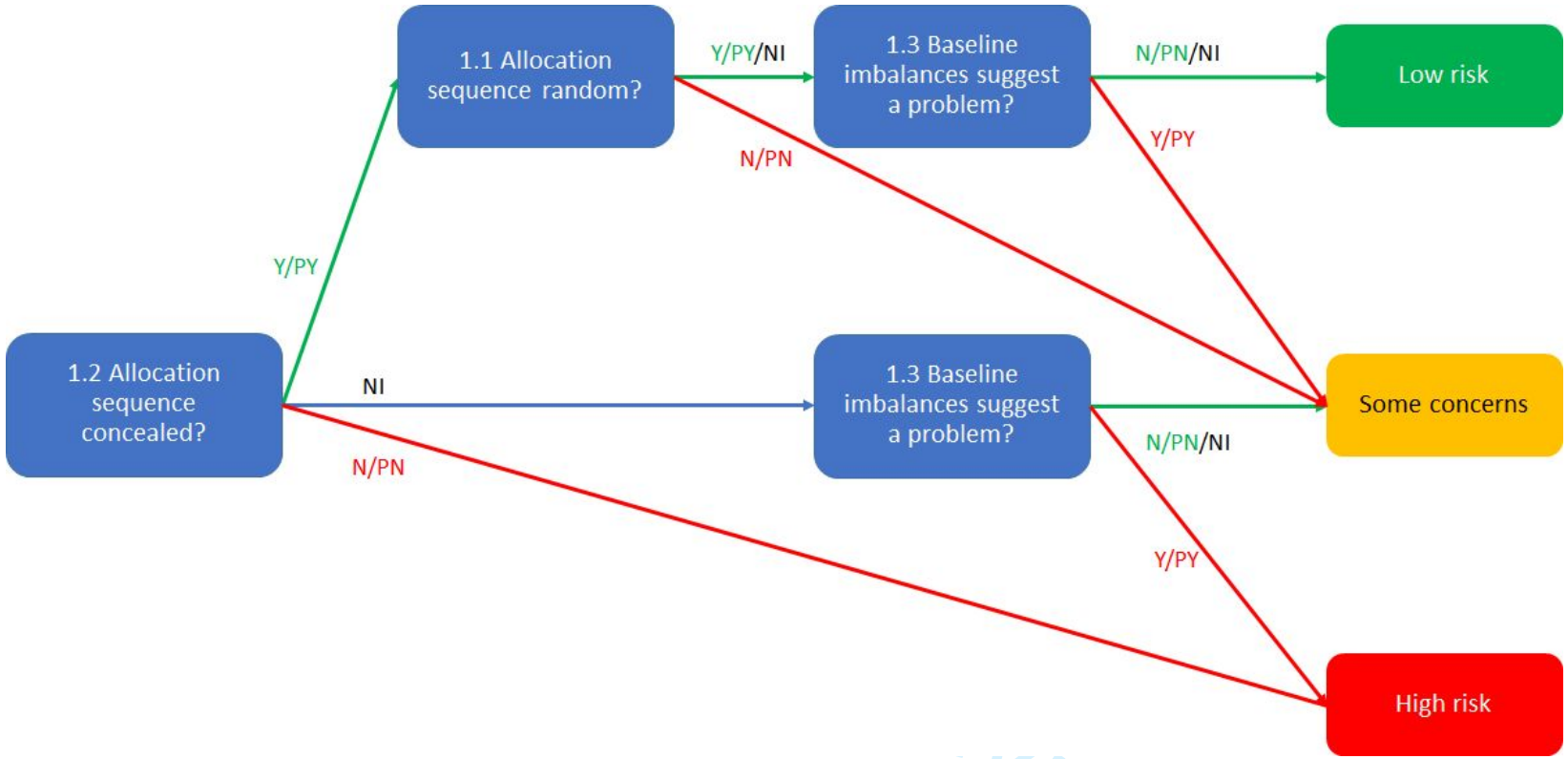
Signalling questions	Elaboration	Response options
1.1 Was the allocation sequence random?	Answer 'Yes' if a random component was used in the sequence generation process. Examples include computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots. Minimization is generally implemented with a random element (at least when the scores are equal), so an allocation sequence that is generated using minimization should generally be considered to be random. Answer 'No' if no random element was used in generating the allocation sequence or the sequence is predictable. Examples include alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method. Answer 'No information' if the only information about randomization methods is a statement that the study is randomized. In some situations a judgement may be made to answer 'Probably no' or 'Probably yes'. For example, , in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomization sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information'. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods.	Y/PY/PN/N /NI
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Answer 'Yes' if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by an external unit or organization, independent of the enrolment personnel (e.g. independent central pharmacy, telephone or internet-based randomization service providers). Answer 'Yes' if envelopes or drug containers were used appropriately. Envelopes should be opaque, sequentially numbered, sealed with a tamper-proof seal and opened only after the envelope has been irreversibly assigned to the participant. Drug containers should be sequentially numbered and of identical appearance, and dispensed or administered only after they have been irreversibly assigned to the participant. This level of detail is rarely provided in reports, and a judgement may be required to justify an answer of 'Probably yes' or 'Probably no'. Answer 'No' if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation.	Y/PY/PN/N /NI

1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	*Note that differences that are compatible with chance do not lead to a risk of bias. A small number of differences identified as 'statistically significant' at the conventional 0.05 threshold should usually be considered to be compatible with chance.* Answer 'No' if no imbalances are apparent or if any observed imbalances are compatible with chance. Answer 'Yes' if there are imbalances that indicate problems with the randomization process, including: (1) substantial differences between intervention group sizes, compared with the intended allocation ratio; or (2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance; or (3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate. Also answer 'Yes' if there are other reasons to suspect that the randomization process was problematic: (4) excessive similarity in baseline characteristics that is not compatible with chance. Answer 'No information' when there is no *useful* baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis). The answer to this question should not influence answers to questions 1.1 or 1.2. For example, if the trial has large baseline imbalances, but authors report adequate randomization methods, questions 1.1 and 1.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to the question 1.3 and reflected in the domain-level risk-of-bias judgement. Trialists may undertake analyses that attempt to deal with flawed randomization by controlling for imbalances in prognostic factors at baseline. To remove the risk of bias caused by problems in the randomization process, it would be necessary to know, and measure, all the prognostic factors that were imbalanced at baseline. It is unlikely that all important prognostic factors are known and measured, so such analyses will at best reduce the risk of bias. If review authors wish to assess the risk of bias in a trial that controlled for baseline imbalances in order to mitigate failures of randomization, the study should be assessed using the ROBINS-I tool.	Y/PY/PN/N/NI

Risk-of-bias judgement	See table below.	Low / High / Some concerns
Optional: What is the predicted direction of bias arising from the randomization process?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

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

Low risk of bias	(i) The allocation sequence was adequately concealed AND (ii.1) Any baseline differences observed between intervention groups appear to be compatible with chance OR (ii.2) There is no information about baseline imbalances AND (iii.1) The allocation sequence was random OR (iii.2) There is no information about whether the allocation sequence was random
Some concerns	(i.1) The allocation sequence was adequately concealed AND (i.2.1) The allocation sequence was not random OR (i.2.2) Baseline differences between intervention groups suggest a problem with the randomization process OR (ii.1) There is no information about concealment of the allocation sequence AND (ii.2) Any baseline differences observed between intervention groups appear to be compatible with chance OR (iii) There is no information to answer any of the signalling questions
High risk of bias	(i) The allocation sequence was not adequately concealed OR (ii.1) There is no information about concealment of the allocation sequence AND (ii.2) Baseline differences between intervention groups suggest a problem with the randomization process



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

Signalling questions	Elaboration	Response options
2.1. Were participants aware of their assigned intervention during the trial?	If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'.	Y/PY/PN/N/NI
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer question 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were 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 experimental context?	For the effect of assignment to intervention, bias in this domain arises only when changes from assigned intervention that are inconsistent with the trial protocol arose because of the experimental context and are likely to affect the outcome. For example, the process of securing informed consent may lead participants subsequently assigned to the comparator group to feel unlucky and therefore seek the experimental intervention, or other interventions that improve their prognosis. Answer 'Yes' or 'Probably yes' if there is evidence that the experimental context led to additional interventions beyond those specified in the protocol, or failure to implement the interventions as intended, or non-adherence by trial participants to their assigned intervention, compared with what would have happened had a trial not been taking place. Answer 'No' or 'Probably no' for changes to intervention that are consistent with the trial protocol, for example cessation of a drug intervention because of acute toxicity or use of additional interventions whose aim is to treat consequences of one of the intended interventions Answer 'No' or 'Probably no' for non-adherence to intervention, provided that this is unrelated to the experimental context. If blinding is compromised because participants report side effects or toxicities that are specific to one of the interventions, answer 'Yes' or 'Probably yes' only if there were changes from assigned intervention that are inconsistent with the trial protocol and arose because of the experimental context. The answer 'No information' may be appropriate, because trialists do not always report whether deviations arose because of the experimental context. However, if such deviations *probably* occurred the answer should be 'Probably yes'.	NA/Y/PY/PN/N/NI
2.4 If Y/PY/NI to 2.3: Were these deviations likely to have affected the outcome?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the experimental context will be important if they affect the outcome, but not otherwise.	NA/Y/PY/PN/N/NI
2.5. If Y/PY to 2.4: Were these deviations from intended intervention balanced between groups?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the experimental context are more likely to lead to bias if they are not balanced between the intervention groups.	NA/Y/PY/PN/N/NI

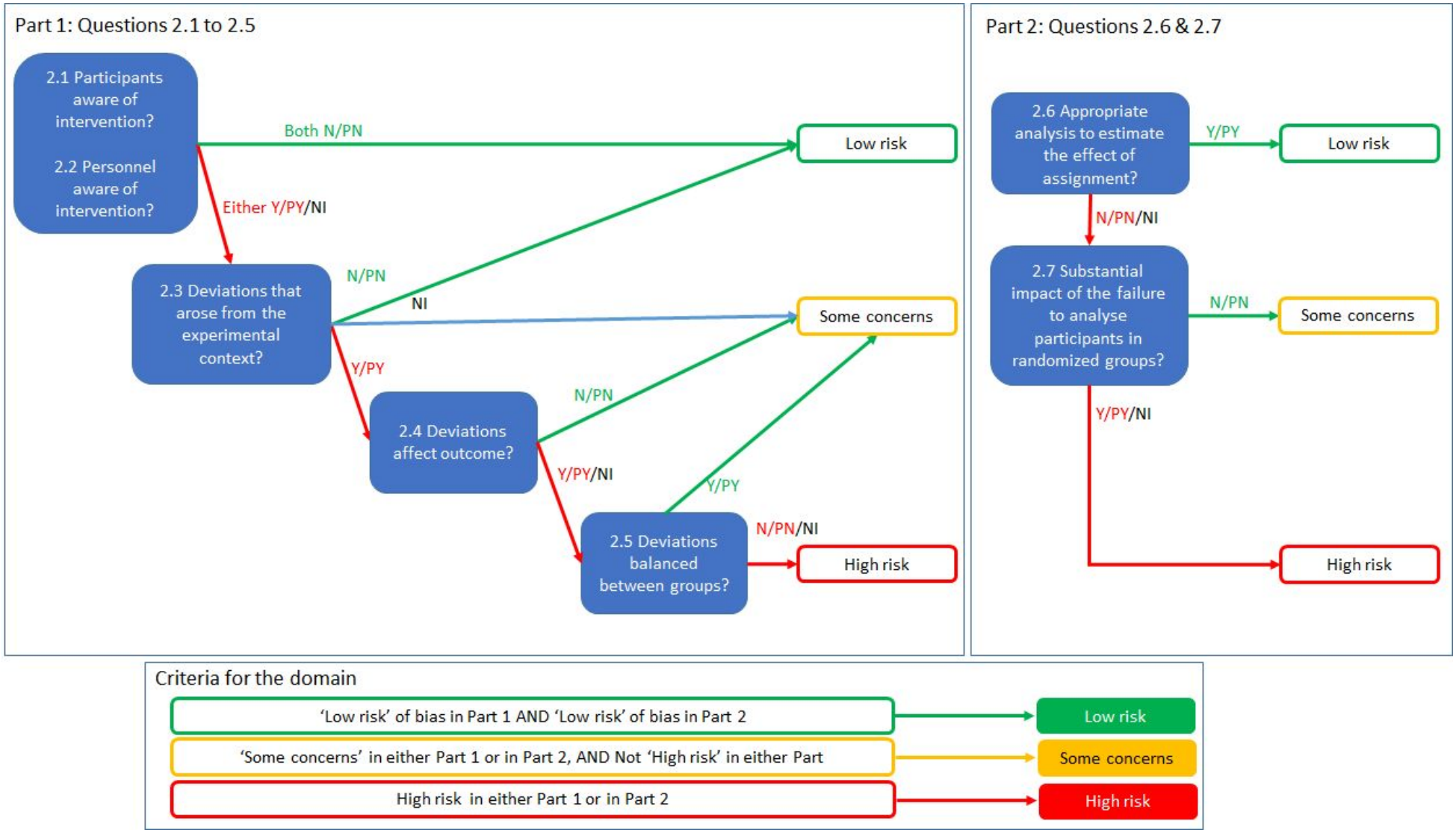
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Both intention-to-treat (ITT) analyses and modified intention-to-treat (mITT) analyses excluding participants with missing outcome data should be considered appropriate. Both naïve 'per-protocol' analyses (excluding trial participants who did not receive their assigned intervention) and 'as treated' analyses (in which trial participants are grouped according to the intervention that they received, rather than according to their assigned intervention) should be considered inappropriate. Analyses excluding eligible trial participants post-randomization should also be considered inappropriate, but post-randomization exclusions of ineligible participants (when eligibility was not confirmed until after randomization, and could not have been influenced by intervention group assignment) can be considered appropriate.	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?	This question addresses whether the number of participants who were analysed in the wrong intervention group, or excluded from the analysis, was sufficient that there could have been a substantial impact on the result. It is not possible to specify a precise rule: there may be potential for substantial impact even if fewer than 5% of participants were analysed in the wrong group or excluded, if the outcome is rare or if exclusions are strongly related to prognostic factors.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See table below.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to deviations from intended interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Algorithm for suggested judgment of risk of bias due to deviations from the intended interventions (effect of assignment to intervention)

Because the domain addresses two somewhat distinct issues, we separate the algorithm into two parts and combine them to reach the judgement.

	Part 1: criteria for questions 2.1 to 2.5	Part 2: criteria for questions 2.6 and 2.7	Criteria for the domain
Low risk of bias	(i) Participants, carers and people delivering the interventions were unaware of intervention groups during the trial OR (ii.1) Participants, carers or people delivering the interventions were aware of intervention groups during the trial AND (ii.2) No deviations from intended intervention arose because of the experimental context.	An appropriate analysis was used to estimate the effect of assignment to intervention	'Low' risk of bias for Part 1 AND 'Low' risk of bias for Part 2
Some concerns	(i) Participants, carers or people delivering the interventions were aware of intervention groups during the trial AND (ii.1) There is no information on whether there were deviations from intended intervention because of the experimental context OR (ii.1.1) There were deviations from intended interventions that arose because of the experimental context AND (ii.1.1.1) These deviations were not likely to have affected the outcome OR (ii.1.1.2) These deviations were balanced between the intervention groups	(i) An appropriate analysis was not used to estimate the effect of assignment to intervention AND (ii) The potential impact (on the estimated effect of intervention) of the failure to analyse participants in the group to which they were randomized was not substantial	''Some concerns' for Part 1 OR ''Some concerns' for Part 2 AND Part 1 not 'High' risk of bias AND Part 2 not 'High' risk of bias

High risk of bias	(i) Participants, carers or people delivering the interventions were aware of intervention groups during the trial AND (ii) There were deviations from intended interventions that arose because of the experimental context AND (iii) These deviations were likely to have affected the outcome AND (iv) These deviations were unbalanced between the intervention groups	(i) An appropriate analysis was not used to estimate the effect of assignment to intervention AND (ii) The potential impact (on the estimated effect of intervention) of the failure to analyse participants in the group to which they were randomized was substantial	'High' risk of bias for Part 1 OR 'High' risk of bias for Part 2


Part 1: Questions 2.1 to 2.5
2.1 Participants aware of intervention?
2.2 Personnel aware of intervention?
Both N/PN
Low risk
Either Y/PY/NI
2.3 Deviations that arose from the experimental context?
N/PN
NI
Some concerns
Y/PY
2.4 Deviations affect outcome?
N/PN
Y/PY/NI
Y/PY
2.5 Deviations balanced between groups?
N/PN/NI
High risk
Part 2: Questions 2.6 & 2.7
2.6 Appropriate analysis to estimate the effect of assignment?
Y/PY
Low risk
N/PN/NI
2.7 Substantial impact of the failure to analyse participants in randomized groups?
N/PN
Some concerns
Y/PY/NI
High risk
Criteria for the domain
'Low risk' of bias in Part 1 AND 'Low risk' of bias in Part 2
Low risk
'Some concerns' in either Part 1 or in Part 2, AND Not 'High risk' in either Part
Some concerns
High risk in either Part 1 or in Part 2
High risk

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

Signalling questions	Elaboration	Response options
2.1. Were participants aware of their assigned intervention during the trial?	If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'.	Y/PY/PN/N/NI
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial.	Y/PY/PN/N/NI
2.3. [If applicable:] If Y/PY/NI to 2.1 or 2.2: Were important non-protocol interventions balanced across intervention groups?	This question is asked only if the preliminary considerations specify that the assessment will address imbalance of important non-protocol interventions between intervention groups. Important non-protocol interventions are the additional interventions or exposures that: (1) are inconsistent with the trial protocol; (2) trial participants might receive with or after starting their assigned intervention; and (3) are prognostic for the outcome. Risk of bias will be higher if there is imbalance in such interventions between the intervention groups.	NA/Y/PY/PN/N/NI
2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?	This question is asked only if the preliminary considerations specify that the assessment will address failures in implementing the intervention that could have affected the outcome. Risk of bias will be higher if the intervention was not implemented as intended by, for example, the health care professionals delivering care. Answer 'No' or 'Probably no' if implementation of the intervention was successful for most participants.	NA/Y/PY/PN/N/NI

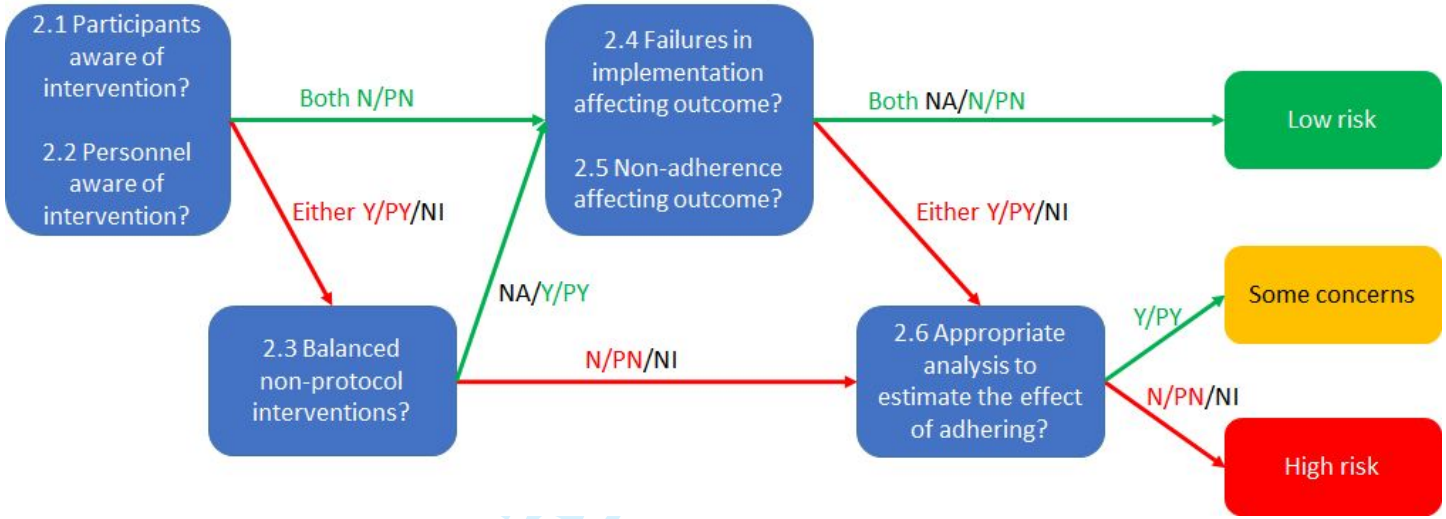
2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants' outcomes?	This question is asked only if the preliminary considerations specify that the assessment will address non-adherence that could have affected participants' outcomes. Non-adherence includes imperfect compliance with a sustained intervention, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention. Consider available information on the proportion of study participants who continued with their assigned intervention throughout follow up, and answer 'Yes' or 'Probably yes' if the proportion who did not adhere is high enough to raise concerns. Answer 'No' for studies of interventions that are administered once, so that imperfect adherence is not possible, and all or most participants received the assigned intervention.	NA/Y/PY/PN/N/NI

2.6. If N/PN/NI to 2.3, or Y/PY/NI to 2.4 or 2.5: Was an appropriate analysis used to estimate the effect of adhering to the intervention?	Both ' naïve 'per-protocol' analyses (excluding trial participants who did not receive their allocated intervention) and 'as treated' analyses (comparing trial participants according to the intervention they actually received) will usually be inappropriate for estimating the effect of adhering to intervention (the 'per-protocol' effect). However, it is possible to use data from a randomized trial to derive an unbiased estimate of the effect of adhering to intervention. Examples of appropriate methods include: (1) instrumental variable analyses to estimate the effect of receiving the assigned intervention in trials in which a single intervention, administered only at baseline and with all-or-nothing adherence, is compared with standard care; and (2) inverse probability weighting to adjust for censoring of participants who cease adherence to their assigned intervention, in trials of sustained treatment strategies. These methods depend on strong assumptions, which should be appropriate and justified if the answer to this question is 'Yes' or 'Probably yes'. It is possible that a paper reports an analysis based on such methods without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information. If an important non-protocol intervention was administered to all participants in one intervention group, adjustments cannot be made to overcome this. Some examples of analysis strategies that would not be appropriate to estimate the effect of adhering to intervention are (i) 'Intention to treat (ITT) analysis', (ii) 'per protocol analysis', (iii) 'as-treated analysis', (iv) 'analysis by treatment received'.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to deviations from intended interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Algorithm for suggested judgment of risk of bias due to deviations from the intended interventions (effect of adhering to intervention)

Low risk of bias	(i.1) Participants, carers and people delivering the interventions were unaware of intervention groups during the trial OR (i.2.1) Participants, carers or people delivering the interventions were aware of intervention groups AND (i.2.2) [If applicable] The important non-protocol interventions were balanced across intervention groups AND (ii) [If applicable] Failures in implementing the intervention could not have affected the outcome AND (iii) [If applicable] Study participants adhered to the assigned intervention regimen
Some concerns	(i.1.1) Participants, carers and people delivering the interventions were unaware of intervention groups during the trial AND (i.1.2.1) [If applicable] Failures in implementing the intervention could have affected the outcome OR (i.1.2.2) [If applicable] Study participants did not adhere to the assigned intervention regimen OR (i.2.1) Participants, carers or people delivering the interventions were aware of intervention groups AND (i.2.2) [If applicable] The important non-protocol interventions were balanced across intervention groups AND (i.2.3.1) [If applicable] Failures in implementing the intervention could have affected the outcome OR (i.2.3.2) [If applicable] Study participants did not adhere to the assigned intervention regimen OR (i.3.1) Participants, carers or people delivering the interventions were aware of intervention groups AND (i.3.2) [If applicable] The important non-protocol interventions were not balanced across intervention groups AND (ii) An appropriate analysis was used to estimate the effect of adhering to the intervention

High risk of bias	(i.1.1) Participants, carers and people delivering the interventions were unaware of intervention groups during the trial AND (i.1.2.1) [If applicable] Failures in implementing the intervention could have affected the outcome OR (i.1.2.2) [If applicable] Study participants did not adhere to the assigned intervention regimen OR (i.2.1) Participants, carers or people delivering the interventions were aware of intervention groups AND (i.2.2) [If applicable] The important non-protocol interventions were balanced across intervention groups AND (i.2.3.1) [If applicable] Failures in implementing the intervention could have affected the outcome OR (i.2.3.2) [If applicable] Study participants did not adhere to the assigned intervention regimen OR (i.3.1) Participants, carers or people delivering the interventions were aware of intervention groups AND (i.3.2) [If applicable] The important non-protocol interventions were not balanced across intervention groups AND (ii) An appropriate analysis was not used to estimate the effect of adhering to the intervention



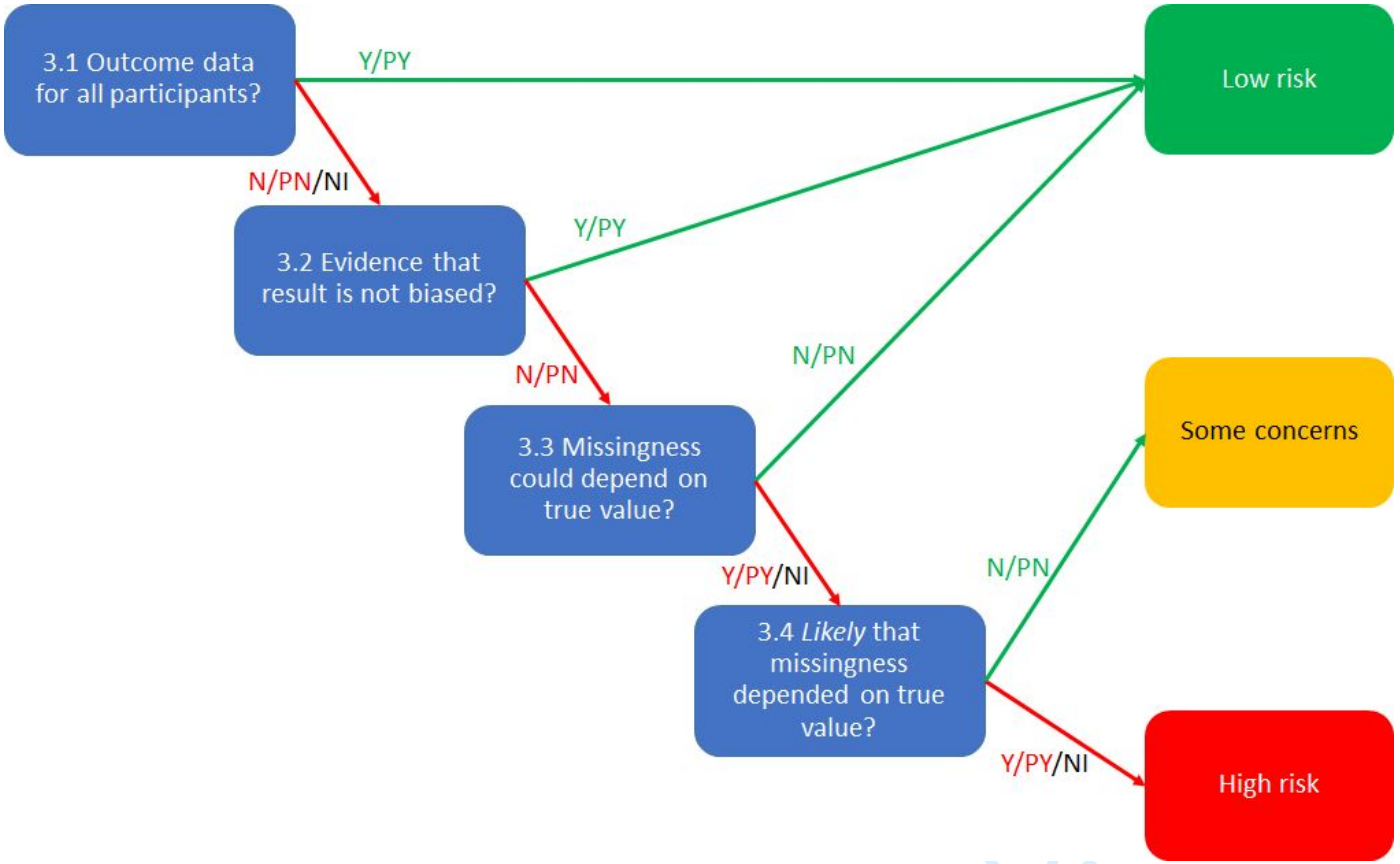
Domain 3: Risk of bias due to missing outcome data

Signalling questions	Elaboration	Response options
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	The appropriate study population for an analysis of the intention to treat effect is all randomized participants. "Nearly all" should be interpreted as that the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention. For continuous outcomes, availability of data from 95% of the participants will often be sufficient. For dichotomous outcomes, the proportion required is directly linked to the risk of the event. If the observed number of events is much greater than the number of participants with missing outcome data, the bias would necessarily be small. Only answer 'No information' if the trial report provides no information about the extent of missing outcome data. This situation will usually lead to a judgement that there is a high risk of bias due to missing outcome data. Note that imputed data should be regarded as missing data, and not considered as 'outcome data' in the context of this question.	Y/PY/PN/N/NI
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as 'last-observation-carried-forward' or via multiple imputation based only on intervention group, should not be assumed to correct for bias due to missing outcome data.	NA/Y/PY/PN/N
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	If loss to follow up, or withdrawal from the study, could be related to participants' health status, then it is possible that missingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection). In time-to-event analyses, participants censored during trial follow-up, for example because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams.	NA/Y/PY/PN/N/NI

3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	This question distinguishes between situations in which (i) missingness in the outcome could depend on its true value (assessed as 'Some concerns') from those in which (ii) it is likely that missingness in the outcome depended on its true value (assessed as 'High risk of bias'). Five reasons for answering 'Yes' are: 1. Differences between intervention groups in the proportions of missing outcome data. If there is a difference between the effects of the experimental and comparator interventions on the outcome, and the missingness in the outcome is influenced by its true value, then the proportions of missing outcome data are likely to differ between intervention groups. Such a difference suggests a risk of bias due to missing outcome data, because the trial result will be sensitive to missingness in the outcome being related to its true value. For time-to-event-data, the analogue is that rates of censoring (loss to follow-up) differ between the intervention groups. 2. Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value; 3. Reported reasons for missing outcome data differ between the intervention groups; 4. The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely. 5. In time-to-event analyses, participants' follow up is censored when they stop or change their assigned intervention, for example because of drug toxicity or, in cancer trials, when participants switch to second-line chemotherapy. 6. Answer 'No' if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See table below.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to missing outcome data?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Suggested judgment of risk of bias due to missing outcome data

Low risk of bias	(i) Outcome data were available for all, or nearly all, randomized participants OR (ii) There is evidence that the result was not biased by missing outcome data OR (iii) Missingness in the outcome could not depend on its true value
Some concerns	(i) Outcome data were not available for all, or nearly all, randomized participants AND (ii) There is not evidence that the result was not biased by missing outcome data AND (iii) Missingness in the outcome could depend on its true value AND (iv) It is not likely that missingness in the outcome depended on its true value
High risk of bias	(i) Outcome data were not available for all, or nearly all, randomized participants AND (ii) There is not evidence that the result was not biased by missing outcome data AND (iii) Missingness in the outcome could depend on its true value AND (iv) It is likely that missingness in the outcome depended on its true value.


3.1 Outcome data for all participants?
Y/PY
Low risk
N/PN/NI
3.2 Evidence that result is not biased?
Y/PY
N/PN
3.3 Missingness could depend on true value?
N/PN
Some concerns
Y/PY/NI
3.4 *Likely* that missingness depended on true value?
N/PN
Y/PY/NI
High risk

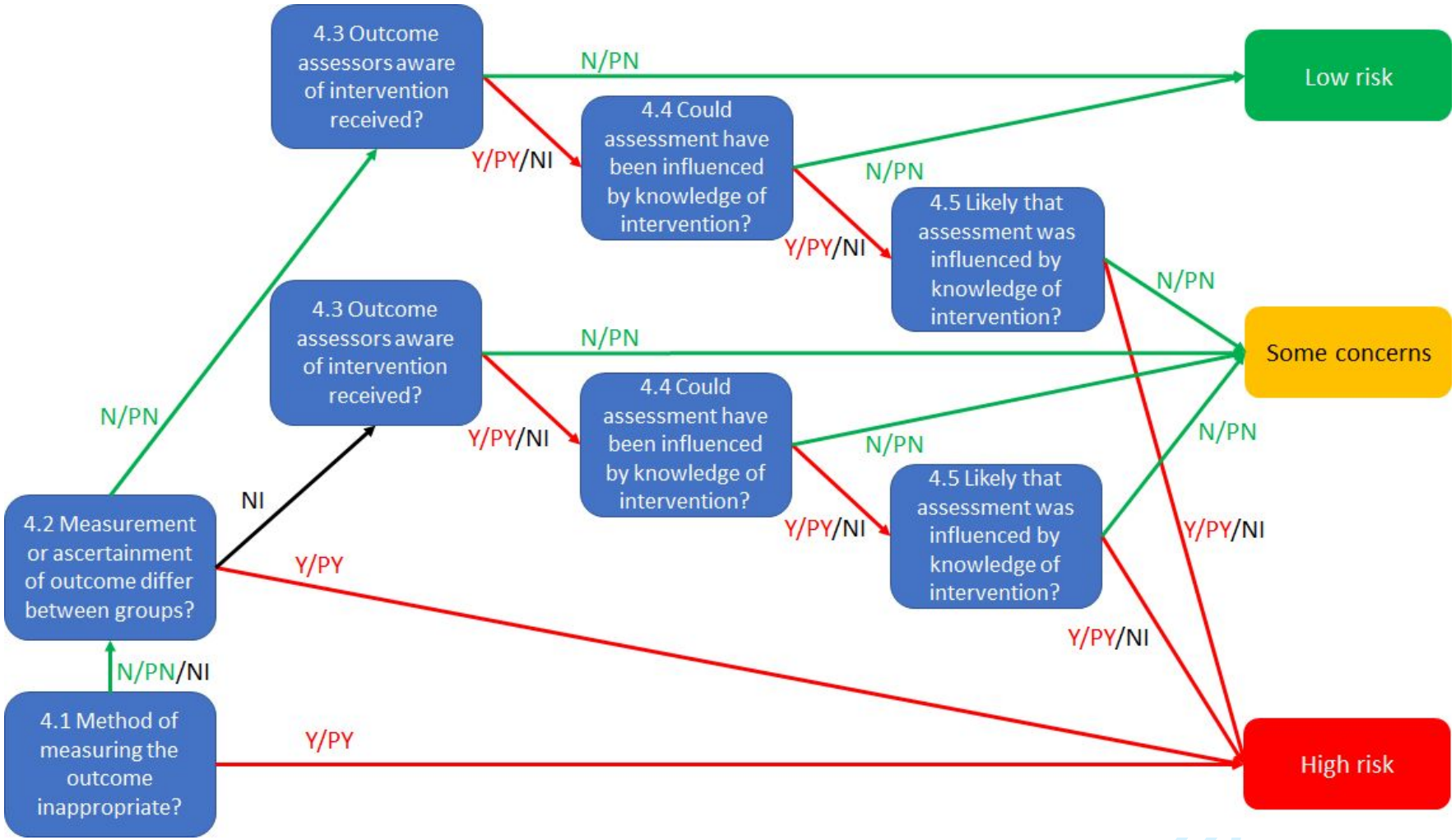
Domain 4: Risk of bias in measurement of the outcome

Signalling questions	Elaboration	Response options
4.1 Was the method of measuring the outcome inappropriate?	This question aims to identify methods of outcome measurement (data collection) that are unsuitable for the outcome they are intended to evaluate. The question *does not* aim to assess whether the choice of outcome being evaluated was sensible (e.g. because it is a surrogate or proxy for the main outcome of interest). In most circumstances, for pre-specified outcomes, the answer to this question will be 'No' or 'Probably no'. Answer 'Yes' or 'Probably yes' if the method of measuring the outcome is inappropriate, for example because: (1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or (2) the measurement instrument has been demonstrated to have poor validity.	Y/PY/PN/N/NI
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of 'diagnostic detection bias' in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified.	Y/PY/PN/N/NI
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Answer 'No' if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant.	Y/PY/PN/N/NI
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer-reported outcomes that do not involve judgement, for example all-cause mortality.	NA/Y/PY/PN/N/NI

4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as 'Some concerns') from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as 'High'). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See table below.	Low / High / Some concerns
Optional: What is the predicted direction of bias in measurement of the outcome?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Suggested judgment of risk of bias in measurement of the outcome

Low risk of bias	(i) The method of measuring the outcome was not inappropriate AND (ii) The measurement or ascertainment of the outcome did not differ between intervention groups AND (iii.1) The outcome assessors were unaware of the intervention received by study participants OR (iii.2) The assessment of the outcome could not have been influenced by knowledge of the intervention received
Some concerns	(i.1) The method of measuring the outcome was not inappropriate AND (i.2) The measurement or ascertainment of the outcome did not differ between intervention groups AND (i.3) The assessment of the outcome could have been influenced by knowledge of the intervention received AND (i.4) It is unlikely that assessment of the outcome was influenced by knowledge of intervention received OR (ii.1) The method of measuring the outcome was not inappropriate AND (ii.2) There is no information on whether the measurement or ascertainment of the outcome could have differed between intervention groups AND (ii.3.1) The outcome assessors were unaware of the intervention received by study participants OR (ii.3.2) The assessment of the outcome could not have been influenced by knowledge of the intervention received
High risk of bias	(i) The method of measuring the outcome was inappropriate OR (ii) The measurement or ascertainment of the outcome could have differed between intervention groups OR (iii) It is likely that assessment of the outcome was influenced by knowledge of the intervention received


4.3 Outcome assessors aware of intervention received?
N/PN
Low risk
Y/PY/NI
4.4 Could assessment have been influenced by knowledge of intervention?
N/PN
Y/PY/NI
4.5 Likely that assessment was influenced by knowledge of intervention?
N/PN
4.3 Outcome assessors aware of intervention received?
N/PN
Some concerns
N/PN
Y/PY/NI
4.4 Could assessment have been influenced by knowledge of intervention?
N/PN
Y/PY/NI
4.5 Likely that assessment was influenced by knowledge of intervention?
N/PN
Y/PY/NI
N/PN
NI
4.2 Measurement or ascertainment of outcome differ between groups?
Y/PY
Y/PY/NI
N/PN/NI
4.1 Method of measuring the outcome inappropriate?
Y/PY
High risk

Domain 5: Risk of bias in selection of the reported result

Signalling questions	Elaboration	Response options
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	If the researchers' pre-specified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalization of the analysis intentions must precede availability of unblinded outcome data to the trial authors. Changes to analysis plans that were made before unblinded outcome data were available, or that were clearly unrelated to the results (e.g. due to a broken machine making data collection impossible) do not raise concerns about bias in selection of the reported result.	Y/PY/PN/N/NI
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...		

5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	A particular outcome domain (i.e. a true state or endpoint of interest) may be **measured** in multiple ways. For example, the domain pain may be measured using multiple scales (e.g. a visual analogue scale and the McGill Pain Questionnaire), each at multiple time points (e.g. 3, 6 and 12 weeks post-treatment). If multiple measurements were made, but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to outcome measurements that are eligible for consideration by the RoB 2 tool user. For example, if only a result using a specific measurement scale is eligible for inclusion in a meta-analysis (e.g. Hamilton Depression Rating Scale), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from a different measurement scale (e.g. Beck Depression Inventory). Answer 'Yes' or 'Probably yes' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements. or There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures). or Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if: Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to	Y/PY/PN/N/NI

	enable an assessment, and there is more than one way in which the outcome domain could have been measured.	

5.3 ... multiple eligible analyses of the data?	A particular outcome measurement may be analysed in multiple ways. Examples include: unadjusted and adjusted models; final value vs change from baseline vs analysis of covariance; transformations of variables; different definitions of composite outcomes (e.g. 'major adverse event'); conversion of continuously scaled outcome to categorical data with different cut-points; different sets of covariates for adjustment; and different strategies for dealing with missing data. Application of multiple methods generates multiple effect estimates for a specific outcome measurement. If multiple estimates are generated but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to analyses that are eligible for consideration by the RoB 2 tool user. For example, if only the result from an analysis of post-intervention values is eligible for inclusion in a meta-analysis (e.g. at 12 weeks after randomization), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from an analysis of changes from baseline. Answer 'Yes' or 'Probably yes' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results arises from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception or vested interest in showing that an experimental intervention is beneficial may be inclined to selectively report analyses that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses. or There is only one possible way in which the outcome measurement can be analysed (hence there is no opportunity to select from multiple analyses). or Analyses are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if:	Y/PY/PN/N/NI

	Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome measurement could have been analysed.	
Risk-of-bias judgement	See table below.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to selection of the reported result?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Suggested judgment of risk of bias in selection of the reported result

Low risk of bias	(i) The data were analysed in accordance with a pre-specified plan that was finalized before unblinded outcome data were available for analysis AND (ii) The result being assessed is unlikely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain AND (iii) Reported outcome data are unlikely to have been selected, on the basis of the results, from multiple analyses of the data
Some concerns	(i.1) The data were not analysed in accordance with a pre-specified plan that was finalized before unblinded outcome data were available for analysis AND (i.2) The result being assessed is unlikely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain AND (i.3) The result being assessed is unlikely to have been selected, on the basis of the results, from multiple analyses of the data OR (ii)There is no information on whether the result being assessed is likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain AND from multiple analyses of the data
High risk of bias	(i) The result being assessed is likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain OR (ii) The result being assessed is likely to have been selected, on the basis of the results, from multiple analyses of the data