Web appendix (published as supplied by the authors)

In this appendix we provide motivation and considerations for assessing the risk of bias for each of the items included in the Cochrane Collaboration's risk of bias assessment tool. More detail is provided in the *Cochrane Handbook for Systematic Reviews of Interventions.*[w1]

Selection bias

The unique strength of randomization is that, if successfully accomplished, it prevents selection bias in allocating interventions to participants. Its success in this respect depends on fulfilling several interrelated processes. A rule for allocating interventions to participants must be specified, based on some chance (random) process. We call this generation of the allocation sequence. Furthermore, steps must be taken to secure strict implementation of that schedule of random assignments by preventing foreknowledge of the forthcoming allocations at the time that participants are recruited to the trial. We refer to this process as concealment of the allocation sequence.

Random sequence generation

While the theoretical basis for the use of randomization is strong, specific empirical evidence on its relationship with bias is limited.[w2] The use of a random component should be sufficient for generation of an allocation sequence. However, we recommend that review authors be confident that the term has been used appropriately before assessing the risk of bias to be low. A simple statement such as 'we randomly allocated' is often insufficient. It is not uncommon for trial authors to use the term 'randomized' even when it is not justified.[w3] If there is doubt, then the adequacy of sequence generation should be regarded as unknown and the risk of bias therefore unclear. Systematic methods, such as alternation, assignment based on date of birth, case record number and date of presentation, are not random although they are sometimes referred to as 'quasi-random'. An important weakness with all systematic methods is that concealing the allocation schedule is usually impossible, which allows foreknowledge of intervention assignment among those recruiting participants to the trial, and biased allocations.

Allocation concealment

Efforts made to generate unpredictable and unbiased sequences are likely to be ineffective if those sequences are not protected by adequate concealment of the allocation sequence from those involved in the enrolment and assignment of participants. Knowledge of the next assignment – for example, from a table of random numbers openly posted on a bulletin board – can cause selective enrolment of participants on the basis of perceived prognosis. Participants who would have been assigned to an intervention deemed to be 'inappropriate' may be rejected. Other participants may be deliberately directed to the 'appropriate' intervention, which can often be accomplished by delaying a participant's entry into the trial until the next appropriate allocation appears. Empirical evidence is strong that concealment of allocation sequence is associated with the effect size. On average, effect estimates are exaggerated by 18% when concealment is rated as inadequate or unclear.[w4]

Strategies for avoiding bias due to inadequate concealment of an allocation sequence in individually randomized trials include central randomization (e.g. with a central randomization office remote from patient recruitment centres), and use of sequentially numbered, identical drug containers.[w5] The use of fixed and known block sizes can, however, undermine allocation concealment.

Performance bias

Performance bias refers to systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest. We recommend assessment of risk of bias arising from lack of blinding of participants and personnel.

Blinding of participants and personnel

After enrolment into the trial, blinding (or masking) of trial participants and personnel may reduce the risk that knowledge of which intervention was received affects outcomes and outcome measurements. Effective blinding can also ensure that the compared groups receive a similar amount of attention, ancillary treatment and diagnostic investigations. Trial reports often describe blinding in broad terms, such as "double blind". This term makes it impossible to know who was blinded.[w6] Such terms are also used very inconsistently,[w7-w9] and should be avoided.

Blinding is not always possible. However, the risk of bias due to lack of blinding must be assessed whether or not blinding is feasible. When blinding is employed, some have suggested that it would be sensible to ask trial participants at the end of the trial to guess which treatment they had been receiving.[w10;w11] Evidence of correct guesses exceeding 50% would seem to suggest that blinding may have been broken, but in fact can simply reflect the patients' experiences in the trial: a good outcome, or a marked side effect, will tend to be more often attributed to an active treatment, and a poor outcome to a placebo.[w12;w13] Therefore evaluations of the success of blinding may need to be interpreted carefully.

When considering the risk of bias from lack of blinding in participants and personnel, it is important to consider first exactly who was and was not blinded; and second the implication of these for the risk of bias in actual outcomes (e.g. due to co-intervention or differential behaviour). Blinding will often need to be addressed for different outcomes separately.

Detection bias

Detection bias refers to systematic differences between groups in how outcomes are determined. We recommend assessment of risk of bias arising from lack of blinding of outcome assessment.

Blinding of outcome assessment

Most outcome assessments can be influenced by lack of blinding, although there are particular risks of bias with more subjective outcomes (e.g. pain or number of days with a common cold). It is therefore important to consider how subjective or objective an outcome is when considering blinding of outcome assessment. Seemingly objective assessments, e.g. doctors assessing the degree of psychological or physical impairment, can be subjective.[w14] In empirical studies, lack of blinding in randomized trials has been shown to be associated with intervention effects that are exaggerated by 9% on average, measured as an odds ratio.[w4] Those studies have dealt with a variety of outcomes, some of which are objective. The estimated effect is more biased, on average, in trials with subjective outcomes.[w15]

Attrition bias

Attrition bias refers to differences between intervention groups in withdrawals from the trial. Missing outcome data, due to attrition (drop-out) during the trial or exclusions from the analysis, raise the possibility that the observed effect estimate is biased. We use the term

incomplete outcome data to refer to both attrition and exclusions. When an individual participant's outcome is not available we shall refer to it as missing.

Incomplete outcome data

Attrition may occur in a clinical trial for several reasons. Participants might withdraw, or be withdrawn, from the trial (e.g. because of an adverse effect); they may fail to attend specific appointments; they may fail to provide specific data; or simply be lost to follow up. Rarely, some data or records may be lost or destroyed. In addition, some participants may be excluded from analysis by the trial investigators, for example if they are enrolled but later found to be ineligible; or if an "as-treated" (or per-protocol) analysis is performed (in which participants are included only if they received the intended intervention in accordance with the protocol).

Assessing the risk of bias due to incomplete outcome data is complex and involves several considerations. Simply observing whether an analysis was described as "intention-to-treat", or determining whether there are more missing data than a pre-specified threshold, are inadequate. Considerations include the extent to which intention-to-treat principles were followed; the proportion of participants with missing outcome data, and whether there is an imbalance between intervention groups; the reasons for missing outcome data; the effect size, and hence the potential impact of missing outcome data on it; and the extent to which bias can be overcome by the review author (e.g. by re-instating excluded participants). Information about these issues should be collected systematically and used to inform a judgement about the risk of bias in the effect estimates presented in the systematic review.

Reporting bias

Reporting bias refers to systematic differences between reported and unreported findings. A widely recognized reporting bias is *publication bias*: that is, the publication or not of whole reports of studies on the basis of their findings.[w16] When assessing the risk of bias in a specific trial, however, the relevant biases arise from selective reporting of particular analyses and findings. We recommend the assessment of bias due to selective reporting of outcomes.

Selective outcome reporting

Within-trial selective reporting bias is a substantial problem in clinical trials.[w17] Selective reporting of outcomes takes several forms, including suppression of all data for an outcome, selection of a biased analysis strategy for the outcome and selection of a biased subset of the data.[w18]

A thorough assessment of the potential for selective reporting bias will be labour intensive. If the protocol for the trial can be located, then outcomes in the protocol and published report can be compared. Alternatively, the trial may have details in a trials registry. If not, particularly for older trials, then outcomes listed in the methods section of an article can be compared with those whose results are reported. If non-significant results are mentioned but not reported adequately (e.g. if only a P value is available), bias in a meta-analysis is likely to occur. Review authors might also consider a small number of key outcomes that are routinely measured in the area in question and report which trials report data on these and which do not. A useful first step is to construct a matrix indicating which outcomes were recorded in which trials, for example with rows as trials and columns as outcomes. Complete and incomplete reporting can also be indicated in such a matrix. A schema for assessing risk of bias due to selective outcome reporting has recently been described.[w19;w20]

The assessment of risk of bias due to selective reporting of outcomes should be made for the trial as a whole, rather than for each outcome. Although it may be clear for a particular trial that some specific outcomes are subject to selective reporting while others are not, we recommend the trial-level approach because it is not practical to list all fully reported outcomes in the risk of bias table. The free-text description should be used to describe the outcomes for which there is particular evidence of selective (or incomplete) reporting. The trial-level judgement provides an assessment of the overall susceptibility of the trial to selective reporting bias.

Other biases

The final domain includes other sources of bias. This domain allows review authors to add one or more specific items that address issues particular to their review, and for which the considerations above do not completely cover anticipated risks of bias. For example, some potential biases are relevant only to particular trial designs (e.g. carry-over effects in cross-over trials and recruitment bias in cluster-randomized trials); and there may be sources of bias that are only found in particular clinical settings (e.g. contamination, a form of performance bias in whereby participants experience some or all of an intervention allocated to a different group). Specific items for this domain should preferably be pre-specified in the review protocol, along with a decision as to whether they will be assessed for trials as a whole, or for individual (or grouped) outcomes within each trial.

Items included in this domain should be direct causes of bias, and should not be (i) sources of heterogeneity (e.g. choice of comparator, length of follow-up), (ii) sources of imprecision or over-precision (e.g. failure to account for clustering); or (iii) quality indicators that are not direct causes of bias (e.g. sample size calculations; ethical approval, source of funding).

References for web appendix

w1. Higgins JPT, Altman DG, editors. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): John Wiley & Sons; 2008. p. 187-241.

w2. Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006; 163(6):493-501.

w3. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* 2009; 10:46.

w4. Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007; 36: 847-857.

w5. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; 359: 614-618.

w6. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002; 136: 254-259.

w7. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001; 285: 2000-2003.

w8. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol* 2005; 58: 1220-1226.

w9. Haahr MT, Hróbjartsson A. Who is blinded in randomised clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006; 3: 360-365.

w10. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004; 328: 432.

w11. Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Contemp Clin Trials* 2005; 26: 25-37.

w12. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 2007; 36: 664-665.

w13. Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. *Lancet* 2010; 375: 1144-1146.

w14. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994; 44: 16-20.

w15. Wood L, Egger M, Gluud LL, Schulz K, Jüni P, Altman DG et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601-605.

w16. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009, Issue 1. Art. No.: MR000006.

w17. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008; 3: e3081.

w18. Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005; 14: 515-524.

w19. Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010; 11: 52.

w20. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010; 340: c365.