

## The Cochrane Collaboration's tool for assessing risk of bias in randomised trials

Julian P T Higgins,<sup>1</sup> Douglas G Altman,<sup>2</sup> Peter C Gøtzsche,<sup>3</sup> Peter Jüni,<sup>4</sup> David Moher,<sup>5,6</sup> Andrew D Oxman,<sup>7</sup> Jelena Savović,<sup>8</sup> Kenneth F Schulz,<sup>9</sup> Laura Weeks,<sup>5</sup> Jonathan A C Sterne,<sup>8</sup> Cochrane Bias Methods Group  
Cochrane Statistical Methods Group

<sup>1</sup>MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 0SR, UK

<sup>2</sup>Centre for Statistics in Medicine, University of Oxford, Oxford, UK

<sup>3</sup>The Nordic Cochrane Centre, Rigshospitalet and University of Copenhagen, Denmark

<sup>4</sup>Institute of Social and Preventive Medicine, University of Bern, Switzerland

<sup>5</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

<sup>6</sup>Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Canada

<sup>7</sup>Preventive and International Health Care Unit, Norwegian Knowledge Centre for the Health Services, Oslo, Norway

<sup>8</sup>Department of Social Medicine, University of Bristol, Bristol, UK

<sup>9</sup>FHI, Research Triangle Park, North Carolina, USA

Correspondence to: J P T Higgins  
julian.higgins@mrc-bsu.cam.ac.uk

Accepted: 22 July 2011

Cite this as: *BMJ* 2011;343:d5928  
doi: 10.1136/bmj.d5928

Flaws in the design, conduct, analysis, and reporting of randomised trials can cause the effect of an intervention to be underestimated or overestimated. The Cochrane Collaboration's tool for assessing risk of bias aims to make the process clearer and more accurate

Randomised trials, and systematic reviews of such trials, provide the most reliable evidence about the effects of healthcare interventions. Provided that there are enough participants, randomisation should ensure that participants in the intervention and comparison groups are similar with respect to both known and unknown prognostic factors. Differences in outcomes of interest between the different groups can then in principle be ascribed to the causal effect of the intervention.<sup>1</sup>

Causal inferences from randomised trials can, however, be undermined by flaws in design, conduct, analyses, and reporting, leading to underestimation or overestimation of the true intervention effect (bias).<sup>2</sup> However, it is usually impossible to know the extent to which biases have affected the results of a particular trial.

Systematic reviews aim to collate and synthesise all studies that meet prespecified eligibility criteria<sup>3</sup> using methods that attempt to minimise bias. To obtain reliable conclusions, review authors must carefully consider the potential limitations of the included studies. The notion of study "quality" is not well defined but relates to the extent to which its design, conduct, analysis, and presentation were appropriate to answer its research question. Many tools for assessing the quality of randomised trials are available, including scales (which score the trials) and checklists (which assess tri-

als without producing a score).<sup>4-7</sup> Until recently, Cochrane reviews used a variety of these tools, mainly checklists.<sup>8</sup> In 2005 the Cochrane Collaboration's methods groups embarked on a new strategy for assessing the quality of randomised trials. In this paper we describe the collaboration's new risk of bias assessment tool, and the process by which it was developed and evaluated.

### Development of risk assessment tool

In May 2005, 16 statisticians, epidemiologists, and review authors attended a three day meeting to develop the new tool. Before the meeting, JPTH and DGA compiled an extensive list of potential sources of bias in clinical trials. The items on the list were divided into seven areas: generation of the allocation sequence; concealment of the allocation sequence; blinding; attrition and exclusions; other generic sources of bias; biases specific to the trial design (such as crossover or cluster randomised trials); and biases that might be specific to a clinical specialty. For each of the seven areas, a nominated meeting participant prepared a review of the empirical evidence, a discussion of specific issues and uncertainties, and a proposed set of criteria for assessing protection from bias as adequate, inadequate, or unclear, supported by examples.

During the meeting decisions were made by informal consensus regarding items that were truly potential biases rather than sources of heterogeneity or imprecision. Potential biases were then divided into domains, and strategies for their assessment were agreed, again by informal consensus, leading to the creation of a new tool for assessing potential for bias. Meeting participants also discussed how to summarise assessments across domains, how to illustrate assessments, and how to incorporate assessments into analyses and conclusions. Minutes of the meeting were transcribed from an audio recording in conjunction with written notes.

After the meeting, pairs of authors developed detailed criteria for each included item in the tool and guidance for assessing the potential for bias. Documents were shared and feedback requested from the whole working group (including six who could not attend the meeting). Several email iterations took place, which also incorporated feedback from presentations of the proposed guidance at various meetings and workshops within the Cochrane Collaboration and from

### SUMMARY POINTS

Systematic reviews should carefully consider the potential limitations of the studies included

The Cochrane Collaboration has developed a new tool for assessing risk of bias in randomised trials

The tool separates a judgment about risk of bias from a description of the support for that judgment, for a series of items covering different domains of bias

## Principles for assessing risk of bias

**1. Do not use quality scales**

Quality scales and resulting scores are not an appropriate way to appraise clinical trials. They tend to combine assessments of aspects of the quality of reporting with aspects of trial conduct, and to assign weights to different items in ways that are difficult to justify. Both theoretical considerations<sup>10</sup> and empirical evidence<sup>11</sup> suggest that associations of different scales with intervention effect estimates are inconsistent and unpredictable

**2. Focus on internal validity**

The internal validity of a study is the extent to which it is free from bias. It is important to separate assessment of internal validity from that of external validity (generalisability or applicability) and precision (the extent to which study results are free from random error). Applicability depends on the purpose for which the study is to be used and is less relevant without internal validity. Precision depends on the number of participants and events in a study. A small trial with low risk of bias may provide very imprecise results, with a wide confidence interval. Conversely, the results of a large trial may be precise (narrow confidence interval) but have a high risk of bias if internal validity is poor

**3. Assess the risk of bias in trial results, not the quality of reporting or methodological problems that are not directly related to risk of bias**

The quality of reporting, such as whether details were described or not, affects the ability of systematic review authors and users of medical research to assess the risk of bias but is not directly related to the risk of bias. Similarly, some aspects of trial conduct, such as obtaining ethical approval or calculating sample size, are not directly related to the risk of bias. Conversely, results of a trial that used the best possible methods may still be at risk of bias. For example, blinding may not be feasible in many non-drug trials, and it would not be reasonable to consider the trial as low quality because of the absence of blinding. Nonetheless, many types of outcome may be influenced by participants' knowledge of the intervention received, and so the trial results for such outcomes may be considered to be at risk of bias because of the absence of blinding, despite this being impossible to achieve

**4. Assessments of risk of bias require judgment**

Assessment of whether a particular aspect of trial conduct renders its results at risk of bias requires both knowledge of the trial methods and a judgment about whether those methods are likely to have led to a risk of bias. We decided that the basis for bias assessments should be made explicit, by recording the aspects of the trial methods on which the judgment was based and then the judgment itself

**5. Choose domains to be assessed based on a combination of theoretical and empirical considerations**

Empirical studies show that particular aspects of trial conduct are associated with bias.<sup>2, 12</sup> However, these studies did not include all potential sources of bias. For example, available evidence does not distinguish between different aspects of blinding (of participants, health professionals, and outcome assessment) and is very limited with regard to how authors dealt with incomplete outcome data. There may also be topic specific and design specific issues that are relevant only to some trials and reviews. For example, in a review containing crossover trials it might be appropriate to assess whether results were at risk of bias because there was an insufficient "washout" period between the two treatment periods

**6. Focus on risk of bias in the data as represented in the review rather than as originally reported**

Some papers may report trial results that are considered as at high risk of bias, for which it may be possible to derive a result at low risk of bias. For example, a paper that inappropriately excluded certain patients from analyses might report the intervention groups and outcomes for these patients, so that the omitted participants can be reinstated

**7. Report outcome specific evaluations of risk of bias**

Some aspects of trial conduct (for example, whether the randomised allocation was concealed at the time the participant was recruited) apply to the trial as a whole. For other aspects, however, the risk of bias is inherently specific to different outcomes within the trial. For example, all cause mortality might be ascertained through linkages to death registries (low risk of bias), while recurrence of cancer might have been assessed by a doctor with knowledge of the allocated intervention (high risk of bias)

pilot work by selected review teams in collaboration with members of the working group. The materials were integrated by the co-leads into comprehensive guidance on the new risk of bias tool. This was published in February 2008 and adopted as the recommended method throughout the Cochrane Collaboration.<sup>9</sup>

**Evaluation phase**

A three stage project to evaluate the tool was initiated in early 2009. A series of focus groups was held in which review authors who had used the tool were asked to reflect on their experiences. Findings from the focus groups were then fed into the design of questionnaires for use in three online surveys of review authors who had used the tool, review authors who had not used the tool (to explore why not), and editorial teams within the collaboration. We held a meeting to discuss the findings from the focus groups and surveys and to consider revisions to the first version of the risk of bias tool. This was attended by

six participants from the 2005 meeting and 17 others, including statisticians, epidemiologists, coordinating editors and other staff of Cochrane review groups, and the editor in chief of the *Cochrane Library*.

**The risk of bias tool**

At the 2005 workshop the participants agreed the seven principles on which the new risk of bias assessment tool was based (box).

The risk of bias tool covers six domains of bias: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. Within each domain, assessments are made for one or more items, which may cover different aspects of the domain, or different outcomes. Table 1 shows the recommended list of items. These are discussed in more detail in the appendix on [bmj.com](http://bmj.com).

For each item in the tool, the assessment of risk of bias is in two parts. The support for judgment provides a succinct free text description or summary of the rel-

**Table 1** | Cochrane Collaboration's tool for assessing risk of bias (adapted from Higgins and Altman<sup>13</sup>)

| Bias domain      | Source of bias                          | Support for judgment   | Review authors' judgment (assess as low, unclear or high risk of bias)   |
|------------------|---|--|--|
| Selection bias   | Random sequence generation              | Describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups  | Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence          |
|                  | Allocation concealment                  | Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen before or during enrolment   | Selection bias (biased allocation to interventions) due to inadequate concealment of allocations before assignment |
| Performance bias | Blinding of participants and personnel* | Describe all measures used, if any, to blind trial participants and researchers from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective   | Performance bias due to knowledge of the allocated interventions by participants and personnel during the study    |
| Detection bias   | Blinding of outcome assessment*         | Describe all measures used, if any, to blind outcome assessment from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective   | Detection bias due to knowledge of the allocated interventions by outcome assessment                               |
| Attrition bias   | Incomplete outcome data*                | Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomised participants), reasons for attrition or exclusions where reported, and any reinclusions in analyses for the review | Attrition bias due to amount, nature, or handling of incomplete outcome data                                       |
| Reporting bias   | Selective reporting                     | State how selective outcome reporting was examined and what was found  | Reporting bias due to selective outcome reporting  |
| Other bias       | Anything else, ideally prespecified     | State any important concerns about bias not covered in the other domains in the tool   | Bias due to problems not covered elsewhere   |

\*Assessments should be made for each main outcome or class of outcomes.

evant trial characteristic on which judgments of risk of bias are based and aims to ensure transparency in how judgments are reached. For example, the item about concealment of the randomised allocation sequence would provide details of what measures were in place, if any, to conceal the sequence. Information for these descriptions will often come from a single published trial report but may be obtained from a mixture of trial reports, protocols, published comments on the trial, and contacts with the investigators. The support for the judgment should provide a summary of known facts, including verbatim quotes where possible. The source of this information should be stated, and when there is no information on which to base a judgment, this should be stated.

The second part of the tool involves assigning a judgment of high, low, or unclear risk of material bias for each item. We define material bias as bias of sufficient magnitude to have a notable effect on the results or con-

clusions of the trial, recognising the subjectivity of any such judgment. Detailed criteria for making judgments about the risk of bias from each of the items in the tool are available in the *Cochrane Handbook*.<sup>13</sup> If insufficient detail is reported of what happened in the trial, the judgment will usually be unclear risk of bias. A judgment of unclear risk should also be made if what happened in the trial is known but the associated risk of bias is unknown—for example, if participants take additional drugs of unknown effectiveness as a result of them being aware of their intervention assignment. We recommend that judgments be made independently by at least two people, with any discrepancies resolved by discussion in the first instance.

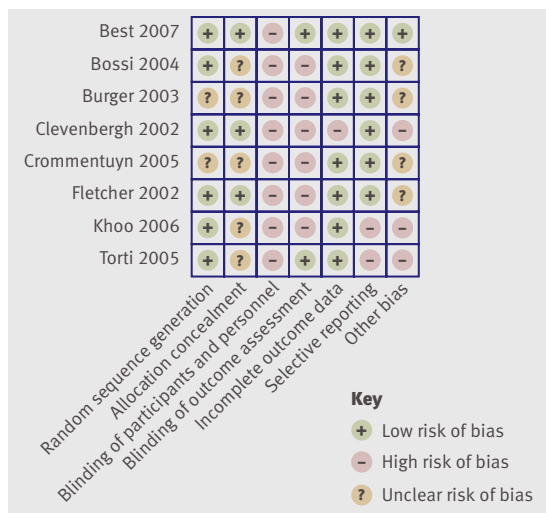
Some of the items in the tool, such as methods for randomisation, require only a single assessment for each trial included in the review. For other items, such as blinding and incomplete outcome data, two or more assessments may be used because they generally need to be made separately for different outcomes (or for the same outcome at different time points). However, we recommend that review authors limit the number of assessments used by grouping outcomes—for example, as subjective or objective for the purposes of assessing blinding of outcome assessment or as “patient reported at 6 months” or “patient reported at 12 months” for assessing risk of bias due to incomplete outcome data.

### Evaluation of initial implementation

The first (2008) version of the tool was slightly different from the one we present here. The 2008 version did not categorise biases by the six domains (selection bias, performance bias, etc); had a single assessment for blinding; and expressed risk of bias in the format “yes,” “no,” or “unclear” (referring to lack of a risk) rather than as low, high, or unclear risk. The 2010 evaluation of the initial version found wide acceptance of the need for the risk of bias tool, with a consensus that it represents an improvement over methods previously recommended by the Collaboration or widely used in systematic reviews.

**Table 2** | Example of risk of bias table from a Cochrane review<sup>14</sup>

| Bias  | Authors' judgment | Support for judgment   |
|---|-------------------|--|
| Random sequence generation (selection bias)                 | Low risk          | Quote: “Randomization was one to one with a block of size 6. The list of randomization was obtained using the SAS procedure plan at the data statistical analysis centre”  |
| Allocation concealment (selection bias)                     | Unclear risk      | The randomisation list was created at the statistical data centre, but further description of allocation is not included   |
| Blinding of participants and researchers (performance bias) | High risk         | Open label   |
| Blinding of outcome assessment (detection bias)             | High risk         | Open label   |
| Incomplete outcome data (attrition bias)                    | Low risk          | Losses to follow-up were disclosed and the analyses were conducted using, firstly, a modified intention to treat analysis in which missing=failures and, secondly, on an observed basis. Although the authors describe an intention to treat analysis, the 139 participants initially randomised were not all included; five were excluded (four withdrew and one had lung cancer diagnosed). This is a reasonable attrition and not expected to affect results. Adequate sample size of 60 per group was achieved |
| Selective reporting (reporting bias)                        | Low risk          | All prespecified outcomes were reported  |
| Other bias  | Unclear risk      | No description of the uptake of the therapeutic drug monitoring recommendations by physicians, which could result in performance bias  |



Example presentation of risk of bias assessments for studies in a Cochrane review of therapeutic monitoring of antiretroviral drugs in people with HIV<sup>14</sup>

Participants in the focus groups noted that the tool took longer to complete than previous methods. Of 187 authors surveyed, 88% took longer than 10 minutes to complete the new tool, 44% longer than 20 minutes, and 7% longer than an hour, but 83% considered the time taken acceptable. There was consensus that classifying items in the tool according to categories of bias (selection bias, performance bias, etc) would help users, so we introduced these. There was also consensus that assessment of blinding should be separated into blinding of participants and health professionals (performance bias) and blinding of outcome assessment (detection bias) and that the phrasing of the judgments about risk should be changed to low, high, and unclear risk. The domains reported to be the most difficult to assess were risk of bias due to incomplete outcome data and selective reporting of outcomes. There was agreement that improved training materials and availability of worked examples would increase the quality and reliability of bias assessments.

**Presentation of assessments**

Results of an assessment of risk of bias can be presented in a table, in which judgments for each item in each trial are presented alongside their descriptive justification. Table 2 presents an example of a risk of bias table for one trial included in a Cochrane review of therapeutic monitoring of antiretrovirals for people with HIV.<sup>14</sup> Risks of bias due to blinding and incomplete outcome data were assessed across all outcomes within each included study,

rather than separately for different outcomes as will be more appropriate in some situations.

Presenting risk of bias tables for every study in a review can be cumbersome, and we suggest that illustrations are used to summarise the judgments in the main systematic review document. The figure provides an example. Here the judgments apply to all meta-analyses in the review. An alternative would be to illustrate the risk of bias for a particular meta-analysis (or for a particular outcome if a statistical synthesis is not undertaken), showing the proportion of information that comes from studies at low, unclear, or high risk of bias for each item in the tool, among studies contributing information to that outcome.

**Summary assessment of risk of bias**

To draw conclusions about the overall risk of bias within or across trials it is necessary to summarise assessments across items in the tool for each outcome within each trial. In doing this, review authors must decide which domains are most important in the context of the review, ideally when writing their protocol. For example, for highly subjective outcomes such as pain, blinding of participants is critical. The way that summary judgments of risk of bias are reached should be explicit and should be informed by empirical evidence of bias when it exists, likely direction of bias, and likely magnitude of bias. Table 3 provides a suggested framework for making summary assessments of the risk of bias for important outcomes within and across trials.

**Assessments of risk of bias and synthesis of results**

Summary assessments of the risk of bias for an outcome within each trial should inform the meta-analysis. The two preferable analytical strategies are to restrict the primary meta-analysis to studies at low risk of bias or to present meta-analyses stratified according to risk of bias. The choice between these strategies should be based on the context of the particular review and the balance between the potential for bias and the loss of precision when studies at high or unclear risk of bias are excluded. Meta-regression can be used to compare results from studies at high and low risk of bias, but such comparisons lack power,<sup>15</sup> and lack of a significant difference should not be interpreted as implying the absence of bias.

A third strategy is to present a meta-analysis of all studies while providing a summary of the risk of bias across studies. However, this runs the risk that bias is downplayed in the discussion and conclusions of a review, so that decisions continue to be based, at least in part, on flawed evidence. This risk could be reduced by incorporating summary assessments into broader, but explicit, measures of the quality of evidence for each important outcome, for example using the GRADE system.<sup>16</sup> This can help to ensure that judgments about the risk of bias, as well as other factors affecting the quality of evidence (such as imprecision, heterogeneity, and publication bias), are considered when interpreting the results of systematic reviews.<sup>17 18</sup>

**Discussion**

Discrepancies between the results of different systematic reviews examining the same question<sup>19 20</sup> and between meta-analyses and subsequent large trials<sup>21</sup> have shown that

**Table 3 | Approach to formulating summary assessments of risk of bias for each important outcome (across domains) within and across trials (adapted from Higgins and Altman<sup>13</sup>)**

| Risk of bias         | Interpretation   | Within a trial                                  | Across trials  |
|----------------------|--|---|--|
| Low risk of bias     | Bias, if present, is unlikely to alter the results seriously | Low risk of bias for all key domains            | Most information is from trials at low risk of bias  |
| Unclear risk of bias | A risk of bias that raises some doubt about the results      | Low or unclear risk of bias for all key domains | Most information is from trials at low or unclear risk of bias   |
| High risk of bias    | Bias may alter the results seriously                         | High risk of bias for one or more key domains   | The proportion of information from trials at high risk of bias is sufficient to affect the interpretation of results |

**bmj.com**

Previous articles in this series

- ▶ Choosing target conditions for test accuracy studies that are relevant to clinical practice (*BMJ* 2011;343:d4684)
- ▶ Brackets (parentheses) in formulas (*BMJ* 2011;343:d570)
- ▶ How to obtain the P value from a confidence interval (*BMJ* 2011;343:d2304)
- ▶ Verification problems in diagnostic accuracy studies (*BMJ* 2011;343:d4770)

the results of meta-analyses can be biased, which may be partly caused by biased results in the trials they include. We believe our risk of bias tool is one of the most comprehensive approaches to assessing the potential for bias in randomised trials included in systematic reviews or meta-analyses. Inclusion of details of trial conduct, on which judgments of risk of bias are based, provides greater transparency than previous approaches, allowing readers to decide whether they agree with the judgments made. There is continuing uncertainty, and great variation in practice, over how to assess potential for bias in specific domains within trials, how to summarise bias assessments across such domains, and how to incorporate bias assessments into meta-analyses.

A recent study has found that the tool takes longer to complete than other tools (the investigators took a mean of 8.8 minutes per person for a single predetermined outcome using our tool compared with 1.5 minutes for a previous rating scale for quality of reporting).<sup>22</sup> The reliability of the tool has not been extensively studied, although the same authors observed that larger effect sizes were observed on average in studies rated as at high risk of bias compared with studies at low risk of bias.<sup>22</sup>

By explicitly incorporating judgments into the tool, we acknowledge that agreements between assessors may not be as high as for some other tools. However, we also explicitly target the risk of bias rather than reported characteristics of the trial. It would be easier to assess whether a drop-out rate exceeds 20% than whether a drop-out rate of 21% introduces an important risk of bias, but there is no guarantee that results from a study with a drop-out rate lower than 20% are at low risk of bias. Preliminary evidence suggests that incomplete outcome data and selective reporting are the most difficult items to assess; kappa measures of agreement of 0.32 (fair) and 0.13 (slight) respectively have been reported for these.<sup>22</sup> It is important that guidance and training materials continue to be developed for all aspects of the tool, but particularly these two.

We hope that widespread adoption and implementation of the risk of bias tool, both within and outside the Cochrane Collaboration, will facilitate improved appraisal of evidence by healthcare decision makers and patients and ultimately lead to better healthcare. Improved understanding of the ways in which flaws in trial conduct may bias their results should also lead to better trials and more reliable evidence. Risk of bias assessments should continue to evolve, taking into account any new empirical evidence and the practical experience of authors of systematic reviews.

**Contributors:** All authors contributed to the drafting and editing of the manuscript. JPTH, DGA, PCG, PJ, DM, ADO, KFS and JACS contributed to the chapter in the *Cochrane Handbook for Systematic Reviews of Interventions* on which the paper is based. JPTH will act as guarantor.

**Development meeting participants (May 2005):** Doug Altman (co-lead), Gerd Antes, Chris Cates, Jon Deeks, Peter Gøtzsche, Julian Higgins (co-lead), Sally Hopewell, Peter Jüni (organising committee), Steff Lewis, Philippa Middleton, David Moher (organising committee), Andy Oxman, Ken Schulz (organising committee), Nandi Siegfried, Jonathan Sterne, Simon Thompson.

**Other contributors to tool development:** Hilda Bastian, Rachelle Buchbinder, Iain Chalmers, Miranda Cumpston, Sally Green, Peter Herbison, Victor Montori, Hannah Rothstein, Georgia Salanti, Guido Schwarzer, Ian Shrier, Jayne Tierney, Ian White and Paula Williamson.

**Evaluation meeting participants (March 2010):** Doug Altman (organising committee), Elaine Beller, Sally Bell-Syer, Chris Cates, Rachel Churchill,

June Cody, Jonathan Cook, Christian Gluud, Julian Higgins (organising committee), Sally Hopewell, Hayley Jones, Peter Jüni, Monica Kjeldstrøm, Toby Lasserson, Allyson Lipp, Lara Maxwell, Joanne McKenzie, Craig Ramsey, Barney Reeves, Jelena Savović (co-lead), Jonathan Sterne (co-lead), David Tovey, Laura Weeks (organising committee).

**Other contributors to tool evaluation:** Isabelle Boutron, David Moher (organising committee), Lucy Turner.

**Funding:** The development and evaluation of the risk of bias tool was funded in part by The Cochrane Collaboration. The views expressed in this article are those of the authors and not necessarily those of The Cochrane Collaboration or its registered entities, committees or working groups. JPTH was also funded by MRC grant number U.1052.00.011. DGA was funded by Cancer Research UK grant number C-5592. DM was funded by a University Research Chair (University of Ottawa). The Canadian Institutes of Health Research provides financial support to the Cochrane Bias Methods Group.

**Competing interests:** None declared

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

- 1 Kleijnen J, Gøtzsche P, Kunz RH, Oxman AD, Chalmers I. So what's so special about randomisation? In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research: on the 25th anniversary of Archie Cochrane's Effectiveness and Efficiency*. BMJ Books, 1997:93-106.
- 2 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;336:601-5.
- 3 Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. BMJ Books, 2001.
- 4 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials—an annotated bibliography of scales and checklists. *Controlled Clin Trials* 1995;12:62-73.
- 5 Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
- 6 West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. *Systems to rate the strength of scientific evidence. Evidence report/technology assessment no 47*. AHRQ publication No 02-E016. Agency for Healthcare Research and Quality, 2002.
- 7 Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *J Clin Epidemiol* 2011;64:79-89.
- 8 Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.
- 9 Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Wiley, 2008.
- 10 Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001;2:463-71.
- 11 Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
- 12 Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;163:493-501.
- 13 Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Wiley, 2008:187-241.
- 14 Kredt T, Van der Walt JS, Siegfried N, Cohen K. Therapeutic drug monitoring of antiretrovirals for people with HIV. *Cochrane Database Syst Rev* 2009;3:CD007268.
- 15 Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23:1663-82.
- 16 Guyatt GH, Oxman AD, Vist GE, Zunj R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
- 17 Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH, et al. Presenting results and "Summary of findings" tables. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Wiley, 2008:335-8.
- 18 Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Interpreting results and drawing conclusions. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Wiley, 2008:359-87.
- 19 Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. *BMJ* 1992;305:913-20.
- 20 Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152-6.
- 21 Leloir J, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.
- 22 Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs SJ, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.