Author's response to reviews

Title: Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies

Authors:

Andreas Lundh (alundh@dadlnet.dk)
Peter C Gotzsche (pcg@cochrane.dk)

Version: 2 Date: 21 February 2008

Author's response to reviews: see over
Author's response to reviews

Title: Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies

Authors:
   Andreas Lundh (alundh@dadlnet.dk)
   Peter C. Gøtzsche (pcg@cochrane.dk)

Version: 2 Date: 21 February 2008

Author's response to reviews: see over
Dear Editor,

We would like to thank the reviewers for their helpful and insightful comments. We have revised the paper in line with their comments and have responded as outlined below.

Reviewer 1

Minor Essential Revisions:
1. The statement in the discussion that “The Jadad scale is the only scale that has been developed using established standards for scales.” Is not an accurate statement. In fact, an additional scale was developed using accepted scale development techniques (Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi list: A criteria list for quality assessment of randomized clinical trials developed by Delphi consensus). Please make a note of this.

   We agree that other tools exist. In their study (A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004;4:22), Katrak et al. found 45 different tools for assessing methodological quality of experimental studies with 24 assigning an overall score. However, the Jadad scale is the only one with reporting of the development process, testing of construct validity, assessment of interrater agreement and for which low scores have been associated with increased estimates of benefit. We have clarified this now.

2. There is an underlying inference in this paper: That reviewers of Cochrane Reviews follow review group guidelines. This inference is likely wrong for many reasons: different questions being asked, heterogeneous reviewer training or expertise, the accumulation of evidence over time, the paucity of evidence on most research on bias, the support given by the various review groups to the reviewer etc. There is empirical evidence that reviewers do not follow the review Group recommendations, generally. Therefore, the jump from review group recommendations to what is actually used by reviewers is large and not warranted. This must be noted with much more clarity in the discussion in this paper.

   We agree that the inference between what is proposed in guidelines and what is actually done in reviews is problematic and have tried to clarify this in our discussion.
3. The authors state in the second full paragraph on page 9, in the second last sentence that “concealment of allocation” is the most important safeguard against bias. I think this statement may be a bit strong. While this item has the most empirical evidence thus far, other items are likely just as important, even though there is only a small amount of empirical evidence on them as of yet. For example, it is very easy to imagine that compliance with the intervention in trials to have an effect on summary treatment effects that is as large or larger than for allocation concealment. That is, if a selection of patients does not take the active treatment medication/intervention (for example) 50% percent of the time, we would expect a rather large influence on the summary effect.

We agree that the sentence is a bit strong, as both lack of concealment of allocation and lack of blinding have been shown empirically to lead to considerable overestimation of the treatment effect. We have rephrased our statement. In relation to compliance we do not regard it as an issue of bias, but rather a reflection of the real-life situation, i.e. what actually happens in the studied patient populations.

Reviewer 2

Minor Essential Revisions:

1. Minor essential revision in the Abstract, last sentence of the Results is missing a word. Should it read: The scales recommended had problems in the individual items and some of the groups recommending components recommended items not related to bias in their quality assessment?

We have corrected this omission.

Reviewer 3

Major compulsory revisions:

1. Please revise the title of the manuscript to reflect the focus of the study (not assessment of risk of bias of studies in reviews per se but a review of guidance given to Cochrane authors by Cochrane review groups about risk of bias assessment). It could be simply reworded to: Guidance on assessing risk of bias in studies in Cochrane reviews.

We agree and have changed the title.

2. Please review the sections of the manuscript where you have referred to the Cochrane Handbook (ref 1 below). There seem to be some errors in the information you have stated (four examples given below):

2a. You state that the Handbook advises “..quality scores should not be used..” (p4 of manuscript) however the Handbook makes the case for using a component approach above a scale approach (for reasons stated in your manuscript) and states that “none of the currently available scales for measuring the validity or quality of trials can be recommended without reservation” and that “if authors or CRGs choose to use such a scale, it must be with caution.” [Thus the Handbook advice is that it is PREFERABLE to use the component approach “ this is a subtle but important point that should be corrected in the manuscript; may go someway to explain why some CRGs have not excluded the scale approach in the past].
We agree that the Handbook is not clear on this topic. However, as the final conclusion in section 6.11 is "Meanwhile, authors should avoid the use of 'quality scores' and undue reliance on detailed quality assessments. It is not supported by empirical evidence, it can be time-consuming and it is potentially misleading" we regard the Handbook as advising against scales. But we have now described the vagueness in the discussion.

2b. You state that the Handbook provides no specific recommendations for incorporating assessments of methodological quality in reviews (p5 of manuscript), and then classify any CRG that “did not provide any recommendations, but referred to the Cochrane Handbook” as giving no advice for using methodological quality assessments of individual studies in reviews..”. The Handbook outlines several methods of incorporating validity assessments into a review (e.g. as threshold criterion, in sensitivity analyses etc) (section 6.10, p86). Please correct this reference to the Handbook and reclassify CRGs for this criterion and modify the results as appropriate.

It is correct that the Handbook in section 6.10 mentions methods for incorporating study validity in reviews (threshold approach, explanatory method, sensitivity analysis, weights, visual plots, subgroup analysis, cumulative meta-analysis and meta-regression). But in our opinion these methods are only mentioned briefly, and apart from reservations towards weighting of studies, the Handbook does not have any preferential endorsement of a specific method. In our opinion, it would not be correct to label the 22 groups without specific recommendations as recommending all of the above methods. We have now clarified and changed our statements to reflect the concerns of the reviewer, however.

2c. You state that the Handbook recommends use of the scale approach (Discussion, p8 para 2). However, the advice is not to score individual validity criteria which are combined to give a numeric summary. Instead the Handbook suggests that an overall summary of the risk of bias (or validity) of the study can be expresses as low, moderate or high risk of bias (depending on the specific criteria assessed and how the bias is likely to affect the results).

According to the method mentioned in section 6.7.1 each study will receive an overall summary of risk of bias based on individual items. While this is not exactly the same as the methods for summing of the scores of individual items into an overall numerical score, we still regard it as a scale approach. Scales combine several features in a single overall value (here low, moderate and high, which is, in fact, a ranking scale that builds on individual components) as opposed to components that only look at individual items.

2d. You state that the Handbook “recommends analyzing all data according to the intention-to-treat principle, but has currently no recommendations on sensitivity analyses related to this item” (Discussion, p9 para 1). See sections 8.4.1 (p112) and 8.10 (p151) of the Handbook for specific recommendations, including guidance on sensitivity analyses in relation to this.

It is correct that section 8.4.1.2 mentions using sensitivity analysis to test different imputation techniques to perform a full intention-to-treat analysis. But this is in relation to different statistical methods for dealing with missing data and minimising bias. It is not related to the validity assessment of studies, e.g. if intention-to-treat
analysis was done for each study. We have now changed the text to avoid the error the reviewer pointed out. 21 groups had recommendations for ITT assessment when assessing the quality of studies, despite the handbook's lack of specific guidance on this. We have now clarified that the Handbook recommends analysing all data according to ITT, but has no recommendation on how to deal with it in relation to validity assessment.

3. Under Methods (p5 para 1), you refer to CRG guidance that “...recommended both scales and components as optional...”. You then state that you classified these groups as recommending scales. It would be more accurate to present these groups separately as “recommending a component and/or scale approach”. In my view these CRGs are different to those advising authors to exclusively use a scale approach.

We agree that there could be a problem in relation to stratifying the data into just two groups (scale vs. components). We initially agreed on these categories to simplify matters. While a sole recommendation is different from an optional recommendation, we wanted the categories to reflect which methods could potentially be used in reviews. The problem is very minor, however, as it only involved two groups. Both groups extracted as optional for a scale approach had direct reference to the Jadad scale and we think it is highly likely that this scale will be used by some of the group's authors. We have changed the discussion to reflect the reviewer's concern.

Minor essential revisions:
1. Please correct the spelling of “adviced” to “advised” throughout the text where necessary (e.g. p5 para 5; p9 para 4 “advises”).

   We have corrected the spelling.

2. Change “Most review groups have it’s to “Most review groups have their” (p3, para 5).

   We have corrected the spelling.

3. Please cite the Issue of The Cochrane Library which was searched for guidance about assessment of methodological quality of studies by CRGs (Methods, p4, para 2). You state that you conducted the study in March 2007 (Methods p4, para 2), however, refer to an outdated version of the Cochrane Handbook for Systematic Reviews of Interventions in your reference list (citation #12). Please update to the version current at the time of conducting the study (i.e. v 4.2.6, updated Sept 2006).

   We have corrected this error.

4. Please present the results in the order that they are described in para 2, p5 [i.e. 1) component vs. scale; 2) areas of quality assessed; 3) type of analytical approach]. Currently #3 is presented in the Results section before #2 (i.e. move para 5 on p5 to the end of the Results section).

   We have changed this and the tables accordingly.
Consider removing reference to specific CRGs in the text and instead refer to “one group recommended” etc.

Whenever we have tried to be discrete in our past papers, about persons, groups, drug companies, or trials, the editors of the journal have invariably asked us to be specific. This directness serves the readers well, as they may then look up more about a particular review group, if they are interested. Also, the review groups’ guidelines have been published, which speaks against not mentioning which groups we are talking about.

Discretionary revisions:
1. The use of the term “slavishly” (pg 7 para 2) has (overly) negative connotations. Please consider rewording.

    We agree and have changed the word to strictly.

2. I am not clear what you mean by “Authors may decide to use a component approach although the group recommends a scale, or vice versa, and some groups are MORE PRODUCTIVE than others?” (what do you mean by more productive?)

    What was meant was that the production of reviews are not equally distributed across groups, and that our percentages are therefore not comparable to those reported in an earlier study that counted reviews, rather than review groups. We have clarified this.

We thank you for the opportunity to revise the paper and hope the paper has improved as a result of the comments. We look forward to your response.

Yours sincerely

Andreas Lundh

Peter. C. Gøtzsche