Author’s response to reviews

Title: Methodological criteria for the assessment of moderators in systematic reviews of Randomised Controlled Trials: a consensus study

Authors:

Tamar Pincus (t.pincus@rhul.ac.uk)
Clare Miles (clare.miles@rhul.ac.uk)
Robert Froud (r.j.froud@qmul.ac.uk)
Martin Underwood (M.Underwood@warwick.ac.uk)
Dawn Carnes (d.carnes@qmul.ac.uk)
Stephanie J.C. Taylor (s.j.c.taylor@qmul.ac.uk)

Version: 2 Date: 30 November 2010

Author’s response to reviews: see over
Response to reviewers comments

Title: Methodological criteria for the assessment of moderators in systematic reviews of Randomised Controlled Trials: a consensus study

We thank the reviewers for their time and their recommended amendments. We have acted on these recommendations and believe that the manuscript is much improved. Specifically:

Reviewer's 1 report:

Reviewer 1 asked for two discretionary revisions:

1. I understand the problem the authors faced in searching relevant papers for inclusion in the traditional databases. They solved the problem adequately. However, these searches could be complemented by forward citation searches in the Web of Science by using the already included papers to find other papers that referred to these.

   We agree with the reviewer that further searches would have improved our preliminary work. Unfortunately, because the key papers are methodological publications, forward citation results in an unmanageable and mostly irrelevant list of citations. We note, however, that members of the panel of experts were invited to supply additional criteria and to cite further publications.

2. The authors suggest that the list of criteria they developed, may now be improved and ameliorated by further discussion between experts. I agree with that, but I think that the criteria also have to be tested empirically. This could be done in my opinion by comparing results of moderator subgroup analysis based on published effect sizes following the criteria to results of moderator subgroup analysis obtained through individual patient data meta-analysis of the same papers. I recognize this is a lot of work, but I think it is worth to mention this way of further research in the discussion part of the paper.

   We thank the reviewer for this excellent idea. Although we are unable to carry out this complex exercise as part of this publication, we have added the proposed study into our discussion section as a recommendation for future research.(page 16)

Reviewer 2 report

1. Major Compulsory Revisions
One issue where greater detail and more explicit recommendations might be required is in the meta analysis of moderator results. At one point they point out the superiority of within trial analyses, but there is a tension here, as within study analyses are likely to almost always be hamstrung by limited power. Presumably, the most likely scenario is the analysis of the same moderators in multiple trials,
all individually underpowered, but their summation in a meta analysis allowing both a more precise estimate of effect, and assessment of consistency. I am still a little unclear on the author’s views here – are they suggesting that such meta analysis of subgroup effects should be the norm, or not? If they are, what are the statistical issues in the meta analysis of interactions? I personally felt that this was a significant omission, but I may have misunderstood their approach here. Subgroup analysis through meta regression is one thing, but what I thought we might be moving towards would be meta analysis of the results of multiple interaction tests from different trials using subgroup analyses. However, that might not be their primary goal. I think it would be helpful to discuss this issue in more detail – for example, are there statistical objections to the meta analysis of the results of interaction analyses from multiple trials?

We recognize that the explicit aims of the consensus study were not sufficiently clear in the previous version of the manuscript. We addressed this in the amended version through the following:

1. We have added the following paragraph to the introduction to clarify the aims of the study:

The aim of this consensus is to provide guidance about inclusion and exclusion of studies into such a review. We recognize that there is a need for guidance on statistical pooling of interactions from sub group analysis in RCTs but we feel this is beyond the scope of the current study.

   (page 4)

2. In the discussion:

   We aimed to develop/facilitate consensus on criteria to inform researchers about inclusion of studies that carried out sub-group analysis in meta-analysis. (page 13)

   And

   We recognize that the criteria proposed are only the first step in helping researchers decide on studies for inclusion in meta-analysis of interactions. We note that even if the criteria from this consensus are applied carefully, the analysis could still be compromised by individual RCTs providing insufficient data, using different statistical models, having different covariates in the model, and measuring the moderator and outcome with different instruments. Thus, there is still a need to develop methodological criteria to guide meta-analysis of moderators, including statistical guidance for the analysis of interactions. (page 15)
3. Minor Essential Revisions
The paper provides a lot of detail about the RAND method, but I think a lot of that could be summarised more effectively, because I don’t think it will be the primary interest of the readers of the paper. They are likely to be systematic reviewers and trial methodologists, so I think a lot of the detail on the RAND could be usefully consigned to an appendix.

We have now removed the details describing the RAND method to an appendix.

The introduction could provide a better rationale for moderator analyses. Acceptability is mentioned as a reason, but there is a good argument that preference trials (despite their weaknesses) would be a useful alternative for assessments of acceptability (as opposed to effectiveness). I would prefer if the introduction was a little more structured, detailing why these analyses are of interest, what the different terms mean (as outlined in their bullet points) and the advantages and disadvantages. The introduction as written is a little unclear and doesn’t serve to set the scene well or provide a good overview of the literature.

We recognize the importance of studying preference trials but it is beyond the scope of the current study. However, we have amended the introduction to provide more structure by adding sub-headings, and by moving the definitions of moderators, mediators and predictors to an earlier section. We have also moved the rationale, so that the headings now provide the following structure: Opening paragraph, definitions, rationale and aims of the study, associated literature (pages 3-4).

I would much prefer that the RAND process was placed in the background, and the focus was on the methodological meaning of the various items that were rated.

We have now placed the RAND details in an appendix.

It would be useful to describe the logic behind each of the ratings in more detail. There are some quite complex ideas in the text, and the authors focus on levels of disagreement without providing any great detail about the meaning of the various criteria.

We have now, on page 6, made explicit reference to Table 1, which provides rationale for each criterion, using, where possible, illustrative quotes from key sources.

This paper has the potential to become a much cited methodological piece, but at present I don’t think it functions well in that regard. I worry that the current presentation would adversely effect the impact of this important work.
As an example, they state that ‘the panel disagreed about the recommendation that sub group analysis is only justifiable in cases where the magnitude of the difference is large enough to support different recommendations for different subgroups’. This is not well explained in either the text or the table.

**We have now provided further detail on the panel’s comments (page 11).**

A similar example is given later, in the discussion of residual variances and power – I really did not understand this issue, and I think readers would be helped if a paper like this explained these in more detail.

**The panel felt uncertain about this criterion, and have now added a comment in the discussion, providing more detail and inviting comments from readers (page 16). Also, Table 1 provides the rationale for this criterion.**

Why is subgroup analysis a priority in musculoskeletal work? I know that reflects the interests of the authors but there are plenty of other areas where there is significant interest, including mental health.

**We agree with the reviewer, and have now cited musculoskeletal and mental health populations as examples in which sub-grouping has become a priority for research, and added that these are only examples (page 3).**

What was meant by the statement ‘improve power in subsequent trials by better selection of target groups for stratification’?

**What we meant was that if there is good reason to suspect that a sub-group (for example, Group A, with an ES of around 0.7) will do much better than the rest of the group (Group B, ES=0.3), subsequent trials stratifying for these groups would improve power, as the overlap between the groups reduce Beta as the distributions move apart.**

At several points, the authors highlight that theoretical and clinical arguments about ‘plausible’ subgroup effects are difficult in the absence of theory. They might also highlight that in many areas, an abundance of theory often means that almost any subgroup hypothesis can be given a plausible ‘theory’, even contradictory ones.

**We agree and have added the following to the discussion:**

We note that ascertaining whether the rationale for sub group analysis through a plausible theory may be particularly difficult in areas were theories are abundant. (page 15)

**• Discretionary Revisions**

The authors might like to think about how their recommendations can be
assessed against, and might supplant, those of Cochrane and CONSORT.

We have concluded the amended manuscript by suggesting future research, including the suggestion that the findings from reviews using these criteria should be compared to the findings from reviews that followed the guidance from the Cochrane handbook and the CONSORT recommendations (page 16).