Reviewer's report

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

Version: 1 Date: 24 October 2010

Reviewer: Peter Bower

Reviewer's report:

This is a well written and useful paper which provides guidance for reviewers in assessing the use of moderators in RCTs. The guidance derives from a RAND consensus process, which seems a useful method to use in a complex area such as this.

• 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?

• 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.

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.

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. 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. 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. 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.

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.

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

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.

• Discretionary Revisions

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

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

'I declare that I have no competing interests'