Reviewer's report

Title: Risk of bias: Power to detect study level moderator effects in meta-analysis

Version: 1 Date: 9 May 2013

Reviewer: Mark Simmonds

Reviewer's report:

While I found this an interesting paper I think there are a number of issues that need addressing.

Major compulsory revisions.

1. Contrary to your claim at the end of the background section there is considerable literature around the limited power and possible bias is estimating moderator effects in meta-analysis (papers by for example, Thompson, Higgins, Berlin). Key among these papers is that the power to detect moderator effects has already been determined algebraically in the absence of heterogeneity (Simmonds & Higgins, Stat. Med. 2007). The findings of that paper have a substantial impact on your work, which you must consider and properly acknowledge.

2. (Background) This section is excessively long. Do not lecture the reader on the history or importance of randomisation, quality assessment or statistical power. This section should cover only issues in meta-analysis particularly important for this paper.

3. The background section focuses on moderator effects that may cause bias (eg low quality), but there are many other “positive” moderator effects (eg. dose-response). Are you interested only in bias-causing effects or all moderator effects? The whole paper needs more clarity on this issue.

4. (Methods) How are moderator effects estimated? Using meta-regression or subgroup analysis? The method must be described as it will affect the power. Generally the statistical analysis methods should be presented in more detail.

5. (Application examples) I fund this section confusing. Are you just describing some practical examples here? If so why the discussion of Monte Carlo simulations? Please clarify this section. Perhaps you could summarise your examples in a table.

6. Given the existence of the paper mentioned above the simulation when there is no heterogeneity is unnecessary and would be better replaced with the correct theoretical power from that paper. A simulation is only needed when there is residual heterogeneity where there is (probably) no algebraic form for the power.

7. The paper mentioned above also shows that the power is a function of the relative sample size in the two groups of studies (or the total sample size for a
50:50 split) so it seems that presenting results in terms of number of trials and sample size in each trial is unnecessary and your figures could be simplified. The “covariate heterogeneity” is a key determinant of power that you should consider. In particular you should present figures for other splits of the data, not just 50:50 and 25:75, noting that it is the total sample size in each group, not the number of trials, that drives the power.

8. I assume you have used a 5% Type I error rate calculate power, but if we are looking for bias a much higher type I error rate would be acceptable, to avoid not spotting bias. Higher error rate means greater power. This should be discussed.

Minor essential revisions

9. (Applications, 1st parag.) Researchers are surely not interested in how many studies are needed to detect a moderator effect. First because it is sample size, not number of studies, that is important, and second because in a meta-analysis we have no control over the number of studies. This should be removed or clarified.

10. Potential for bias is a major issue in estimating moderator effects in meta-analysis. This should be discussed in the background or discussion section.

Discretionary revisions

11. (Methods) I am concerned that a moderator effect of only 0.4 at most is too small. If an effect is small it may not be clinically or practically meaningful. Larger effects are of greater interest. If possible give results for larger effect sizes (eg 1 SD)

12. (Methods) Heterogeneity is better quantified using $I^2$ rather than $\tau^2$ for ease of interpretation and to maintain heterogeneity levels as numbers of trials change. If possible perform the simulation in terms of $I^2$.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests