Reviewer's report

Title: Mean difference or standardised mean difference: which is more generalizable and powerful in meta-analyses of continuous outcomes? A meta-epidemiological study

Version: 3 Date: 10 January 2014

Reviewer: Matthew Page

Reviewer's report:

This is a very interesting paper that provides empirical evidence and methodological guidance on an issue that is important to systematic reviewers who need to deal with continuous outcomes. Overall the study is well designed and reported. I have a couple of suggestions to improve the quality of the manuscript.

Minor essential revisions

1. In both the ‘Background’ and ‘Discussion’ section, the only limitation of the SMD that is mentioned is that it may have limited interpretability given that it is in units of standard deviation (SD). However, another important limitation that is important to discuss is that, as summarised by Thorlund et al. “…if the variability or heterogeneity in the severity of patients” condition (and thus the variability in scores on the chosen outcome) varies between trials, their SDs will also vary. As a result, trials that enrol a heterogeneous group of patients will yield smaller SMDs than trials enrolling less heterogeneous patients, even though the unstandardized mean difference estimates—and thus the absolute estimate of the magnitude of treatment effect—may be similar across all trials” (Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. Research Synthesis Methods 2011;2(3):188-203.)

2. In paragraph 1 of the section subtitled ‘Methods, Selection of meta-analyses’, the second eligibility criterion states that eligible meta-analyses had to include at least three studies. It is not clear if this “minimum of three studies” criterion was arbitrarily selected, or based on some pre-specified rationale?

3. In paragraph 2 of the section subtitled ‘Discussion, Interpretability’, the authors state that a limitation of the study is that not all of the analysed outcomes were necessarily a primary outcome of the review. It is not exactly clear why this is a limitation. Can the authors elaborate on why they think their generalizability and power results would differ depending on whether the meta-analysis outcome was primary or secondary?

4. An issue worth commenting on in the Discussion section is the extent to which the generalizability and power results found are likely to differ depending on whether the continuous outcome was an objective outcome (e.g. blood pressure, weight) versus a subjective outcome (e.g. scores on a depression scale). That is,
particular types of outcomes are likely to be more prone to unstable measurement, and therefore may be more heterogeneous across different studies than other, more reliably measured outcomes.

Discretionary revisions
1. In the abstract it is stated that “1068 reviews were included”. I think it is clearer to state that “1068 meta-analyses were included”

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

**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