Reviewer's report

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

Version: 1 Date: 16 May 2013

Reviewer: Joseph Lau

Reviewer's report:

The paper by Hempel et al. in essence asks the question of whether many of the current meta-epidemiological studies to identify potential moderator effects (study quality features) are sufficiently powered to take on this task. The authors constructed a Monte Carlo simulation study using a wide range of parameters as well as on a cross section of 200 randomly selected RCTs published in 2006. In addition, the authors performed post-hoc power calculations on four previously published meta-epidemiologic datasets.

The conclusion that “negative findings (from meta-epidemiological research) should not be considered evidence of lack of effects” is not surprising and I doubt that most researchers in this area would think otherwise. Even (hypothetically) if meta-epidemiological studies show no significant difference between randomized studies and non-randomized studies, I would not conclude that therefore randomization is unnecessary. Thus, I see this paper mainly contributes to the methodological discussion by illustrating how difficult (infeasible or perhaps nonsensical) it is to carry out this type of meta-epidemiological research. To be able to identify potential moderator effect, one needs a very large number of trials that is not present in the current published research on this topic. The number of trials needed further increases in the presence of heterogeneity.

While the simulation covered a wide range of values (number of subjects in trials, number of trials in meta-analyses, different tau-squared values, and effect sizes), it did not take into the account of potential different magnitude of moderator effects in heterogeneous clinical settings of the meta-analyses. For example, Schultz et al reported in JAMA in 1995 after examining 33 obstetrics meta-analyses that trials with inadequate method of treatment allocation resulted in 41% large effects, trials with unclear method of treatment allocation have 30% larger effects, and trials not double blind have 17% larger effects. As you noted, these observations were not consistently observed by others examining similar questions across other medical domains. While blinding could be readily achieved in drug trials, it is unachievable most of the time in surgical trials. Similarly there are differences of moderator effects for objective versus subjective outcomes. Further, even if this observation in a cross-section of time is true, could it change over time?

Thus, moderator effect may vary by clinical conditions, the type of outcomes reported, and other factors. While some of the moderators may have the same effect across all clinical settings, some may be quite different. Modeling this
heterogeneity into the simulation will likely require more trials and meta-analyses for meta-epidemiological studies.

My take-home message from this simulation study is that to demonstrate a moderator effect using the current approach of meta-epidemiological studies, one needs a very large number (perhaps infeasibly so) of trials from many meta-analyses. Even if we are able to have adequately powered meta-epidemiological studies in the future, will these results be useful to inform systematic review practice? The results of these analyses are findings of various degrees of association and/or ratio of odds ratio and the like. The challenge of using this information is given a specific meta-analysis, how would one apply and interpret moderator effects from meta-epidemiological studies? Unless we know for certain that a specific factor is truly universal across all clinical conditions, study designs, etc. It would be difficult to interpret and apply these results. Should one be adjusting the effect size of meta-analysis using meta-epidemiological studies?

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I have no interest to declare.