Reviewer's report

Title: Quantifying, displaying and accounting for heterogeneity in meta-analysis

Version: 4 Date: 1 January 2011

Reviewer: Gerta Rücker

Reviewer's report:

- Description of the manuscript -
This is an empirical study on 18 IPD meta-analyses, focussing on comparing various measures of between-study heterogeneity that were previously proposed in the literature. The authors do not present new methods in this article. They conclude that reviewers should try to explain heterogeneity, if found, by using pre-specified subgroup analyses and sensitivity analyses, and demonstrate this using two examples at hand, the two meta-analyses with largest heterogeneity in their sample.

Though the article lacks new methods and also the conclusions are not really original or surprising, the authors’ results deserve publication as providing (i) empirical results of their study and (ii) R code for measuring heterogeneity following the work of DerSimonian and Kacker.

- Major Compulsory Revisions -
Page 8/ Table 2: The authors conduct a subgroup analysis for the Cisplatin trials. In Table 2, however, instead of presenting results of a full subgroup analysis (with Q-within and Q-between), they only show that heterogeneity is reduced by omitting the other 3 trials. What does this explain? The same holds for the second example (Cervix data).

Page 9/ Figure 4: It is said that for the second example, heterogeneity is very much a product of one single trial, the CeCa trial. Can you discuss whether this trial has some special features that distinguish it from the other trials and explain the difference? That is, as the variable ‘trial’ is not an explanatory variable in itself, we would rather expect other properties (population, design, etc.) for explanation.

- Minor Essential Revisions -
page 5, after equation (3): Please add that Q(tau^2) follows a Chi-square distribution under the null hypothesis of no heterogeneity.

page 9 and also page 11: Wrong spelling of the name Viechtbauer.

page 11, second paragraph, last but one line: Delete a superfluous ‘the’.

Figures 5 and 6: In my reproduction of the manuscript, these figures had the size of a postage stamp, and thus I was not able to interpret them.

- Discretionary Revisions -
I very much appreciate that the authors discuss (on page 9/10) a point I have also been concerned with: In the presence of heterogeneity, the random effects model is often schematically applied without further consideration where the heterogeneity comes from, though sometimes heterogeneity is caused by a small-study effect, in which case the random effects model is a bad idea. I should like to point out that this is discussed in a recent letter (Rücker et al., Stat Med. 2010 Dec 10;29(28):2963-5).

Also interesting in this context – and also for the first paragraph of the Conclusions section – are approaches of adjusting treatment effect estimates for small-study effects (see the Papers by Moreno et al.; also see Rücker et al. Biostatistics. 2011 Jan;12(1):122-42.)

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.