Reviewer's report

Title: Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no-treatment, placebo and active intervention

Version: 1 Date: 31 August 2008

Reviewer: Joseph Lau

Reviewer's report:

The authors seek to quantify the relative contributions of spontaneous improvement, effect of placebo, and effect of active interventions in change from baseline in randomized trials.

The title needs to be changed to reflect the types of trials analyzed. Even though the authors stated the limitations in the discussions, these are mostly trials with subjective measurement endpoints and clearly cannot be generalized to other settings. The nature of the treatments in these trials should be made clearer so that readers can judge what settings are the findings of this study applicable.

It is unclear to me what is exactly the question addressed by the authors in this paper. They appear to ask the question of what is the spontaneous improvement in each arm of a 3-arm trial (no treatment, placebo, and active treatment) compared with their respective baseline but they also performed indirect comparisons of the results among the 3 arms. While the question of change from baseline within each arm is valid, it has little usefulness in interpreting trial results. The effect of a treatment is typically assessed as the difference of the mean difference between comparison groups. Thus whatever spontaneous change that occurs in the active treatment group is also cancelled out by change in the no treatment or placebo arm. There are methodological/statistical problems when the authors used the individual arms results to perform the indirect comparisons, which the authors acknowledged in the discussion. I would suggest that the authors create a simple diagram depicting the 3-arms of a trial and what question(s) they are asking.

In this study, the authors updated a previous Cochrane review and combined 3 arms trials from their previous Cochrane review. The authors stated that it was not straightforward to do the analyses and ended up combining the 3 treatment arms separately comparing the post-treatment values with the values at baseline. This was not a meta-analysis of a specific treatment effect but an attempt at deriving an estimate of a study design effect. The method of combing individual arms across studies ignores the correlations of the treatment arms within the study and the principle of meta-analysis of combining the effect sizes (differences between arms and not just the measurements of the individual arms). Combining just the results of the same treatment arms across studies, rather than the difference of effect size, ignores potentially differences in the baseline values that may exist, and could be influenced by the well known Simpson’s paradox.
problem. The authors stated that they find it reassuring that the overall effect of placebo was -0.28, which was similar to the previous estimate of -0.24. Hover, having a number close to another does not necessarily mean that the answer is correct. In consultation with a statistician with in depth meta-analysis expertise, I was advised that this problem could be handled by analysis of variance or regression methods.

While I appreciate that in this data set “no treatment” has the smallest magnitude of spontaneous improvement, and placebo has larger effect, and active treatment the largest effect, I am not certain whether the actual estimates from the meta-analyses coming up with estimated has much generalizability beyond the data analyzed. The data set is from a relative small sample of 43 trials across 9 topics which were quite heterogeneous. Thus, sampling from another data set might easily yield different results. The authors reports that “thus, on average, the relative contributions . . . were 20% and 25%, respectively.” It is unclear to me how these relative contributions were calculated. If I take -0.19 change for spontaneous improvement relative to -1.03 for active treatment, the relative contribution might be ~20%. For -0.47 change for placebo, shouldn’t the relative contribution be about ~50%?

Did the authors incorporated methodological quality of the studies in their analysis as a possible explanation of heterogeneity?

There are many more (58) active treatment arms than no treatment or placebo arms (43 trials each). How were the extra treatment arms handled in the analysis?

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

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I declare that I have no competing interests.