Reviewer's report

Title: Can we rely on the best trial? A comparison of individual trials and systematic reviews

Version: 1 Date: 28 November 2008

Reviewer: Ian Shrier

Reviewer's report:

The authors ask a very important problem and this is clearly stated in the second paragraph of the background: "Hence if no review is currently available then clinicians or guideline writers must decide whether to rely on the best single study or invest the considerable effort involved in doing a systematic review". This is a serious problem for busy (and even not so busy) clinicians that the evidence-based movement has seemed to ignore.

The authors later outline the nuances: "The evidence process then consists of two stages: (i) find the best and largest trial and then (ii) appraise and (if appropriate) apply the trial results. The first step is not straightforward, but is only worthwhile if the second step will usually provide a sufficiently good answer compared with an up-to-date systematic review. Therefore, as an initial exploration of this process, we decided to examine how well and how frequently the single “best” trial might answer the clinical question.

I have several philosophical problems with the authors’ proposed approach. First, the vast majority of the work goes into the first phase. One spends a lot of time finding all the trials that are relevant to the clinical question. The actual extraction of numbers and statistical analysis is quite easy to do. So why not add the relatively quick step of extracting the numbers?

Second, the authors correctly stated the question as finding the "best" trial. However, their analysis is limited to the largest trial. There are two sub-issues here. (1) I believe the authors are trying to use the "most precise trial". Interestingly, this is NOT ALWAYS the largest trial because precision is determined by variance and sample size is only one factor. For example, with dichotomous data, the proportion of subjects with events also the variance and the effect is different depending on whether one is measuring risk difference, relative risk or odds ratios. So, a recommendation to use the largest trial would be a recommendation to use a less precise estimate in some contexts. I do not think that this limitation invalidates the approach, but the authors should address these issues within the analyses. (2) The largest trial is not necessarily the best trial. Even if we assume that all studies are exquisitely done, the "best" trial in the context of a clinical question is one that provides an unbiased answer to the question. If the population of the large trial is different from the clinical context, it may be internally valid but still biased for the question being asked (and so may be the meta-analysis because the large trial is given so much weight!). And of
course, large trials are not all exquisitely done and some may be not be internally valid. So in the end, one still has to evaluate the strengths and weaknesses of each trial in order to appropriately apply evidence. Again, this represents the bulk of the work in any meta-analysis so why stop short of the final relatively quick step?

I am also not quite sure how to interpret the analyses provided. The largest trial will almost always provide the greatest weight of evidence. The percentage of weight will depend on how many trials are included, total number of subjects, whether random or fixed effects models were used (effect dependent on heterogeneity of data - the authors did not appear to extract this data which is essential to a proper interpretation because the results of large trials will always be closer to the results of a fixed effects meta-analysis compared to a random effects meta-analysis), whether one adjusts for biases in the studies using advanced techniques, etc. All of these factors need to be included when assessing whether a "largest" trial approach is useful. For example, the largest trial when there is only one trial will always be exactly equal to the meta-analysis.

Table 1: I would guess that for most clinical problems, most of the trials are small. Indeed, the median of the "sample size in largest trial" in Table 1 is only 183 subjects. It would seem to me that conclusions based on one trial with only 183 subjects is more likely to be different from a complete meta-analysis on the topic compared to when the largest trial has several thousand subjects. This could be shown in Figure 1 by producing different plots for different ranges of subjects in the largest trial (e.g. 4 plots on the same page, my preference), or different symbols for the different ranges of subjects. I would suggest categories of <100 subjects, 100-500 subjects, 501-1000 subjects, and >1000 subjects as a start based on my personal beliefs but the choice would certainly depend partly on the data and partly on the authors' own experience.

Finally, the conclusions are not based on the data. The authors have appropriately outlined the limitations of the largest trial approach in the discussion, but then seem to ignore this in the conclusions where they treat the large trial as the gold standard. Authors of systematic reviews need to check agreement between every study and the overall conclusion to determine if there are outliers. The reasons for these outliers need to be explored including data entry errors, different clinical context, biases affecting internal validity, etc. There is nothing in this paper that suggests such an analysis should be limited to large trials only, or that discrepancies between large trials and results of meta-analyses are any more important than discrepancies between results of smaller trials and the results of meta-analyses. In other words, the conclusions are simply general principles that should be followed for all studies in a meta-analysis and while re-iterating this is helpful to improve the conduct of meta-analyses, it does not seem an appropriate conclusion for this manuscript.

In summary, the authors have undertaken a great amount of work trying to address a critical question of great relevance to clinicians. I think the data needs more exploration, issues related to random or fixed effects estimates and variance need to be addressed, and the conclusions need to be rephrased.
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 have no competing interests