Author's response to reviews

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

Authors:

Paul P Glasziou (paul.glasziou@dphpc.ox.ac.uk)
Sasha Shepperd (Sasha.Shepperd@dphpc.ox.ac.uk)
Jon Brassey (jon@tripdatabase.com)

Version: 2 Date: 17 February 2009

Author's response to reviews: see over
Dear Editors,

We would like to thank the reviewers for their helpful comments. All appeared to agree that the paper addressed an important question, and that the study made substantial progress in addressing that question. I particularly note Ian Shrier’s comment that sums up the paper well:
“...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.” His comments align with the other reviewers: “This study is clear and well written, making an important contribution to knowledge in this area” (Hopewell) and “The study is of interest. The data are clearly presented.” (Gluud).

The reviewers also make a number of useful suggestions, which we have tried to address, and are detailed below. A summary of the main changes we have made are:

1. To have added a additional Figure which illustrates the two approaches we are comparing: systematic review versus a rapid “best study” review. This new Figure also helps us to spell out that we are only addressing part of the latter approach.
2. We have also modified Figure 1 (now Figure 2) in line with Reviewer 1 and Reviewer 3’s suggestions: this gives an analysis for when the index trial is > and < 50%.

Below we have spelled out in more detail our responses to the suggestions by the reviewers. We have made modifications for most of the comments. However, a few of the comments suggested the reviewers had misinterpreted our aim, but we hope the new Figure 1 and the accompanying new text will have clarified this. We hope you are happy with the revised version, and look forward to hearing from you soon,

Regards
Paul Glasziou
Reviewer 1: Ian Shrier

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.

1. 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?

RESPONSE: We don’t agree that “relatively quick step of extracting the numbers” – the appraisal of every potential trial, the extraction process, the analytic checks, etc take a good deal of time. (We asked several reviewers about the percent of time spent on searching versus the other elements of a systematic review and these ranged from 5% to 30%).

CHANGE: we have added the new Figure 1 to make clearer the contrast between the processes.

2. 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.

(A) 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.

RESPONSE: We agree – we are suggesting “most precise trial”, which is the one with the largest weight in the systematic review.

CHANGE: We have changed the term throughout to the “index trial” and more clearly defined this.

(B) 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?
RESPONSE: We agree that the index trial may be flawed, but not that this means ALL trials need to be critically appraised.

CHANGE: We have added a new Figure to make clearer the process we are examining, and how it contrasts with the systematic review. The Figure (and its discussion) show that if the “index trial” has important flaws, then other trials may need to be examined.

3. 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.

RESPONSE: We agree these factors needed to be consider, but have left that to these to the individual reviewers decision making. These will influence the review but not the index trial.

4. 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.

RESPONSE: Good idea!

CHANGE: We have redone the Figure. We decide to split by weight rather than size though, as number of subjects does not always reflect the power of the study or the relative weight.

5. 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.

RESPONSE: We did not state that the largest trial is the Gold Standard. But we believe if should have greater attention than small trials. For example, a review in 2004 made a serious error by getting the outcome direction the wrong way around (new ref 11 Murray et al. Interactive Health Communication Applications for people with chronic disease. CDSR). The reviewers would have been alerted to this if they had compared their result to the conclusions of the index trial (which did have the correct conclusion).

CHANGE: We have added this example and reference to the Discussion.

Reviewer 2: Sally Hopewell

Thank you for asking me to review this manuscript which aims to evaluate how often, and under what circumstances, using trials with the greatest statistical weight arrives at similar conclusions to the full meta-analysis in a random sample of Cochrane reviews. This study is clear and well written, making an important contribution to knowledge in this area, as such my comments are minor:

Page 6; para 1: The single largest trial was selected from the meta-analysis reporting the primary outcome in each Cochrane review. Was there ever a situation where there were two trials with very similar weight but with very different effect estimates, and if so, what did you do?

RESPONSE: We looked through the forest plots to see if there are 2 trials with similar weights but very different effect estimates. Among all the there are 4 where the weights are similar and the estimates of effect differ in that they are on different sides of the line of no effect but the confidence intervals overlap - so the answer is no.

NO CHANGE.

Page 6; para 4: You mention that it was not possible to tell which trial had the greatest weight in the meta-analysis for four of the 200 randomly selected Cochrane reviews. Is it possible to give an explanation for why this was not possible.

RESPONSE: There were 4 reviews with multiple trials but no meta-analysis, so we could not tell which had the largest weight.

CHANGE: Clarified wording in the Results.

Page 7; para 2: The average number of trials included in the meta-analyses was 7.3 with the largest trials contributing on average 51% of the statistical weight to the summary estimate from the whole meta-analysis. There was agreement between the single largest trial and overall meta-analysis 81% of the time. If would be interesting to know how this percentage agreement varied depending on the number trials included in a meta-analysis. For example, if a meta-analysis contained only a small number of trials compared to one containing a large number of trials (where the largest trial would in theory have less weight).

RESPONSE: We agree it would be interesting, and have partly addressed this with the new Figure 2, but have not sub-grouped the tabulations as we think the numbers are too small, and would add complexity rather than clarity.
Page 9; para 3: One of the limitations of this study is how to find the largest single trial to answer a specific clinical questions without doing a comprehensive search of the literature. The authors knowledge that this is a problem and beyond the scope of this study. However, is it possible to draw on guidance from other areas to help guide clinicians who might be wanting to apply this approach to their own clinical setting.

RESPONSE/CHANGE: We agree it is beyond the scope, but have made that clearer in the new Figure 1, and by more detail in the Discussion.
Reviewer 3: Christian Gluud

Reviewer's report:
Glasziou and co-workers have compared the estimated intervention effects and P-values of meta-analyses in Cochrane systematic reviews to the estimated intervention effects and P-values of the trial contributing most weight to the meta-analysis. There seems to be reasonably good agreement between the trials and the meta-analyses. The study is of interest. The data are clearly presented. I have the following suggestions for the authors.

* Major Compulsory Revisions

1. The comparison seems biased towards finding agreement as the authors seem to compare 167 evaluable meta-analyses, of which 35 only contained a single trial. Thereby they may skew the observed association towards agreement. A more proper analysis seems to be between the trial contributing most weight to the meta-analysis and the meta-analyses containing two or more trials. Sensitivity analyses could then be between the trial contributing most weight to the meta-analysis and the meta-analyses containing three or more trials; four or more trials; and five or more trials.

RESPONSE/CHANGE: We hope the new Figure 1 has made the question we are addressing clearer – the comparison of the Rapid “best trial” review with the meta-analysis. We clarified that Figure 2 only uses the meta-analyses with 2 or more trials. We have also added some sensitivity analysis on the effect of trial weight to the revised Figure 2.

2. The authors use 'best' (e.g., title) and 'largest' interchangeably about their 'index' trial. This seems confusing. I suggest they use the same term every time, e.g., a more correct term, i.e., the trial carrying most weight.

RESPONSE: We agree this is confusing terminology.

CHANGE: Throughout we now use the term “index trial” for the trial in the meta-analysis with greatest weight. We also use the term “most precise trial” in Figure 1, but this only becomes the “index trial” if it were of sufficient quality to be included in the systematic review.

3. The background can be shortened. If I remember correctly The Women's Health Initiative trial provided evidence that hormone replacement carried health harm. That was projected by a meta-analysis published about 10 years earlier in the BMJ. So maybe this example is not the most well selected?

RESPONSE: We agree that the review predicted the results of the WHI. But for a clinician needing a rapid answer, the WHI is sufficient, and is an example that will be familiar to readers.

4. I have difficulties accepting that the authors have chosen the wrong 'gold standard'. Both randomized clinical trials and systematic reviews of randomized clinical trials may reach wrong conclusions due to systematic errors ('bias'); random errors ('play of chance'); and design errors (e.g., wrong comparator, etc.). Accordingly, a comparison of the trial carrying most weight with the result of a meta-analysis of the same trial plus
additional trials has to consider all error mechanisms. I think the authors come ‘too easy about this fact’. So, how many of the trials had adequate methodology and hence low risk of bias among the 'heavy' trials and among all trials in the meta-analysis? How well was the association between the findings of the 'heavy' trials and their meta-analysis when only those trials carrying less than 50% of the weight was included? Which outcomes were assessed? This is especially relevant for the discussion on the influence of bias on the results (Wood, BMJ, 2008).

**RESPONSE/CHANGE:** The new Figure 1 clarifies the need to check for bias in the “most precise trial”, and potentially reject it. The revised Figure 2 provides the analysis for index trials with less than 50% of the weight.

6. How does the authors define an adequately powered trial (p. 9)? Please see recent provocative discussion from Gordon Guyat's group that we do not need that much large trials any more after we have got systematic reviews.

**RESPONSE:** We agree we did not define adequate.

**CHANGE:** We have added the wording “(that is the confidence intervals exclude values that would change the clinical decision)”.

7. I have problems with the proposed 'poor man's systematic review' approach: if you can't find a systematic review then try to find a 'large' trial and use this for guidance. The problem is that when we identify one large trial (however defined) we do not know what we have missed. I agree that it takes time to do systematic reviews - but compare this to the time it takes to do a trial, then systematic reviewing comes out as the most cost-effective winner.

**RESPONSE:** We certainly agree that systematic reviews are cheap and good value for effort, and wish there were more of them! However, if I have a critically ill patient TODAY I cannot wait 6 months for the systematic review. Our process is more about the “urgent man's” review than the “poor man's” review.

**CHANGE:** We have added a discussion and new reference (13) about the cost-effectiveness of doing systematic reviews compared to new trials. We have also added to the Introduction “Clinicians therefore often do a rapid “best trial” review process which will take a few hours or less – over a hundred-fold reduction in effort. The two processes are now more clearly contrasted in Figure 1.”

8. I think the authors need to discuss information size considerations in meta-analyses, e.g., as estimated in trial sequential analyses of cumulative meta-analyses.

**RESPONSE:** We could not see how this was relevant to our article as we are addressing the answer to a question at one point in time.

9. The authors should reconsider their statistical analyses. By using meta-analyses that have been carried out they are able to identify the trial having most weight. Would it not then be more relevant to assess that trials sensitivity and specificity, their predictive value for the result of the meta-analysis (PVpos and PVneg) as well as the likelihood ratio of positive and negative trials for the ultimate result of its meta-analysis?

**RESPONSE:** While interesting for meta-analysts, we don’t see that as the correct framing for the clinical issue we are addressing.
* Minor Essential Revisions
1. P.2 One option could be in stead of is.
RESPONSE/CHANGED: “could be” replaces “is”

2. The authors should use trials and not studies, if it is trials that are meta-analyzed.
RESPONSE/CHANGE: Agree – we have changed to “trials” in 8 places (but left the more generic “studies” in the Introduction, until we focus on trials.

3. The authors should use 'outcome measure' in stead of 'endpoint'
RESPONSE/CHANGE: Agree. We may replaced endpoint.

4. P. 10. 'lower quality' should become 'lower methodological quality (and hence increased risk of systematic errors ('bias')).
RESPONSE/CHANGE: Agree – and changed.