Reviewer’s report

Title: The Orthopaedic Trauma Literature: An Evaluation of Statistically Significant Findings in Orthopaedic Trauma Randomized Trials

Version: 1 Date: 6 November 2007

Reviewer: asbjørn hrobjartsson

Reviewer’s report:

General

The manuscript describes a systematic review of randomised clinical trials that report statistically significant treatment effects within orthopaedic traumatology published between 1995 and 2004. The review quantified the magnitude of treatment effects and the sample sizes with the aim of 1) examining whether the statistically significant effects represented a clinically meaningful treatment effect; and 2) examining whether those small studies with large effects overestimate the true treatment effect. Binary outcomes were summarised as relative risk reduction (RRR), and continuous outcomes were summarised as standardised mean difference (effect size, ES). A clinically meaningful continuous effect was defined as ES > 0.8 and a clinically meaningful binary outcome as RRR < 0.5. 76 trials were identified with 122 continuous and 62 binary outcomes. Almost 2 in 3 of the statistically significant continuous outcomes were clinically meaningful, i.e. ES > 0.8, and almost 1 in 2 of the binary outcomes were of a size considered clinically meaningful, i.e. RRR < 0.5. Large treatment effect (RRR) was strongly correlated to few outcome events. The authors conclude that ‘Readers should interpret the results of such small trials with these issues in mind’.

In general, the systematic review comes across as conducted appropriately. The aims could be formulated clearer, but the methods applied seem valid, and the reporting is clear. The results are not surprising, but the subject matter has not previously been studies in orthopaedic surgery. I think the manuscript could be improved stylistically by shortening, and by focusing on why this subject matter is (especially) important to orthopaedic surgery. I also suggest reflecting more explicitly on the risk of bias due to excluded trials and the problem of defining a unison threshold when an effect becomes ‘clinically relevant’ across different settings.

-----------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

None

-----------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of
1) It is a problem that 52 trials were excluded because of lacking data. This constitutes a high proportion of trials (52/(52+76)=41%), and it is likely that such trials reported lower treatment effects than the trials included. Did the trials excluded because of lacking data differ from the other trials? If the excluded trials with lacking data tended to have lower effect sizes how would that effect the result of the review?

2) I am somewhat uncomfortable with the wording of the aims of the study. The first aim was to examine 'whether the statistically significant effects represented a clinically meaningful treatment effect'. What constitute a meaningful clinical effect depend very much on the clinical situation and the nature of the outcome. Furthermore, even if it was possible to uncontroversially compare RR or effect size in one setting with a RR in another setting there is no consensus of an appropriate threshold. I think this predicament could be more clearly pointed out in the discussion and also emphasised in the aims section by inserting a parenthesis with the operationalisation of 'clinically meaningful effects'.

The second aim was to examine 'whether those small studies with large effects may be overestimates of the true treatment effect'. It seems too loose to include in an aims section whether large effects 'may be overestimates'. I suggest that the aim is reformulated, e.g. to study whether there was an association between large effects and few number of events.

3) The review identifies 76 trials with 62 binary outcomes. Were there any trials with more than one binary outcome? The review also identified 122 continuous outcomes. Again, there must have been some trials that reported more than one outcome. How were such correlated outcomes handled? If correlated outcomes simply were regarded as independent that should be made explicit and commented on.

4) Trials with binary outcomes were analysed for the association between total number of events and treatment effect. Why was a similar analysis not conducted for trials with continuous effects, i.e. a study of the association between effect size and some measure of study precision (e.g. 1/standard error)?

Discretionary Revisions (which the author can choose to ignore)

5) I suggest shortening the introduction substantially. The two examples given (peri-operative beta-blocker and leukaemia chemotherapy) appear redundant and are repeated in the discussion.

6) In the introduction I miss a section that clearly states why this problem is important for orthopaedic surgery.

7) I think it is unfortunate that the authors have chosen to limit their sample by orthopaedic subspecialty instead of by time. It would have been easier to
generalise results if the sample studies was based on newer orthopaedic trials (e.g. from the last 2-4 years) or from a random sample of trials from the last ten years. Still, it would be surprising if results were radically different.

8) The methods section contain quite a bit of explanatory information (e.g. what is an effect size and how did Cohen develop his rather arbitrary cut points of 0.2, 0.5 and 0.8). I think this type of information should be shortened substantially, and would fit better in the discussion in a section on how the results could be interpreted.

9) I miss a section on definitions. It could help a reader if the cut points for ‘clinically meaningful effects’ were made more explicit.

10) The summary of main findings in the beginning of the introduction comes across as too general. I suggest rephrasing the section.

11) The discussion does not discuss in depth the possible mechanisms behind the found relation between total number of events and effect size. The manuscript focus on one aspect, i.e. early termination of the trial, however other equally relevant reasons could be listed, e.g. publication bias, outcome reporting bias, or general lower methodological quality is small trials run by trialists with insufficient methodological training.

12) I think the readability of the manuscript could be improved by shortening the discussion substantially, especially the section ‘relevant literature review’.

What next?: Accept after minor essential revisions

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

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests