Reviewer's report

Title: No Short-Cut in Assessing Trial Quality: A Case Study.

Version: 1 Date: 2 October 2007

Reviewer: Peter Gøtzsche

Reviewer's report:

A number of the author’s observations are very important. Some more work is needed, however, in my opinion. The paper needs shortening, should become more focused, and should present numerical data that can convince the readers that what is being discussed is serious, see below.

Abstract
Suggest mentioning which trial is being discussed, for example by writing ‘Burke et al., BMJ 1991’.

There should not be several full stops in a row.

P2, line 1: Poor quality is not a conundrum, it is a problem.

The author should discuss that many tools for assessing trial quality contain items that have nothing to do with bias; such items are therefore not relevant. What we try to do by assessing trials is to address the risk of bias.

Para 3: Don’t write ‘we’ when there is only one author.

P3, para 2: The author writes that two or more independent reviewers usually assess trials for possible bias, but here it would seem at first that the current trial is only assessed by the author. However, below, the author describes that a class of postgraduate medical students also looked at this trial, but I would like to see a little information as to whether they agreed or disagreed with the author. I also suggest the help of these people gets mentioned in the acknowledgment section.

Para 3: We need a search strategy for identifying systematic reviews of treatment of otitis media. It is a bit confusing to start a sentence by ‘A review based …’; the author should explain that this was an inclusion criterion for the selected reviews.

It is highly recommendable to discuss a trial report with a class of medical students or PhD students as the author did, as this often leads to the unravelling of important problems that are difficult to spot by only one observer.

P4, para 5: Here we learn that another group of students also evaluated the trial; this information should be moved upwards to p3, and it should be discussed how disagreements were resolved.

P5, 1st para: The author needs to quote the systematic reviews rather than
including them in an appendix. I find it problematic that the submitted manuscript plus appendices take up 41 pages although in small print as in a printed journal article. In my opinion, all this material needs to be condensed, preferably into only one article that focuses on the main issues and leave out the minor ones.

Para 2: I am confused by the fact that all these students evaluated the trial by Burke et al. as being of good quality, whereas we now are facing a lot of criticism. I assume that not all of this criticism was generated solely by the author. Did the students contribute in two stages, e.g. first evaluating the study according to a checklist, and then looking at it in more detail? This needs to be clarified.

Para 3: Again, there should not be so many pages of text about this trial as it is then likely that few people would read the account. Much better to condense it.

The author lists 30 problems and calls them major problems. I have not come across a trial that had 30 major problems, and I suggest that the author focuses on the most relevant ones, as the readers may lose sight of what really matters as regards the validity of this trial. Some of the problems can also be combined, e.g. baseline differences occur more than once. The problems should be grouped, e.g. those pertaining to the randomization, those pertaining to missing data, erroneous data, erroneous data analyses.

I suggest that the author incorporates some of the material in the appendices in the article, focus on fewer points, and document these points very thoroughly so there is no doubt in the reader’s mind that this trial is unreliable. For example, the author can document exactly how much baseline values in crying differed, and perhaps show the results using various analyses to make up for missing values or other problems in the original trial report. This way, the criticism could be very productive and would avoid being seen by some as a bit sterile (it is possible to list a large number of problems with most trials, but it is important to distinguish between problems that are important for the estimates of effects and harms).

P8, bottom: Again, it should be shown in this paper how Burke et al.’s study distorted some of the quoted systematic reviews. It would be instructive and very easy to in- or exclude Burke et al. in the Cochrane review, for example, which would give the readers some impression of the importance of this particular trial for the meta-analyses in the field.

P9, para 4: Here we are told about total confounding in terms of age and otorrhoea; this is an example showing that it is difficult for the readers to fully understand all of the arguments in this paper. Having seen no data on this, it is not possible to judge the validity of the argument.

P11, para 2: The author should consider mentioning that the Cochrane Handbook for Systematic Reviews actually disrecommends using scales for quality assessment and favours assessment of individual items such as concealment of allocation. Also, there is a possibility of correcting the Cochrane review in this area, since Cochrane reviews are updated when new trials appear,
or when important criticism has been raised. The author should consider noting this, also because Cochrane reviews are becoming the main source of information for clinicians; there are currently more than 3,000 Cochrane reviews in the Cochrane Library.

I very much agree with the author that critical assessment of trial reports is a very daunting task that needs a lot of attention to detail. It is actually a type of detective work that aims at discovering in particular what the authors of the trial reports may have wished to conceal.

The study by Burke et al. is now called a fatally flawed study. This is probably correct, but I think readers would like to see more convincing evidence in this paper in terms of numbers that can convince them that this was the case. The argumentation has all been done in words, and not in numbers. This should be corrected.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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