Author’s response to reviews

Title: Development and Evaluation of a Quality Score for Abstracts

Authors:
Dr Antje Timmer (antje.timmer@klinik.uni-regensburg.de)
Lloyd R Sutherland (lsutherl@acs.ucalgary.ca)
Robert J Hilsden (rhilsden@acs.ucalgary.ca)

Version: 4 Date: 18 Jan 2003

Comments to the Reviewers

Reviewer 1, Dr. Scherer: Thank for your advice to publish.

Reviewer 2, Dr. Moher: We agree that the issues you brought up are important in themselves.

1. Concealment of allocation, other methodology issues
   We have included a statement relating to the study by Schulz, as suggested (discussion section).
   Also, as suggested, we have included parts of our comments 7 (p. 4/5) and 9 (p.5) into the manuscript.

2. Weighting of randomization vs., e.g., exact p-values
   As pointed out in our last letter, randomization was in fact weighted more heavily as there are several points relating to randomization, while there is, for example, only one item for p-values. We have now clarified this point by adding a sentence into the results section (p 7). May we also draw your attention to no 16 of our reference list, the instrument by Cho and Bero which was used as a basis for developing our instrument. Cho and Bero tried several weightings and did not find this to change the psychometric properties of their instrument.

3. Confidence intervals or exact p-values
   We are not alone in attaching importance to the reporting of confidence intervals. For example, in about all issues of the Users’Guides to the Literature series from JAMA, measures of the effect of chance, in particular measures of the precision of estimates are considered relevant items for the appraisal of studies in evidence based medicine. Cho and Bero originally suggested to ask for confidence intervals or exact p-values. During the development phase of our instrument, we discussed to make confidence intervals a mandatory requirement. However, because the use of confidence intervals was very rare in the pilot sample, we decided not to change the original item. I admit to not being perfectly happy with this decision. On the other hand, the inclusion of too many criteria which are practically never met (concealment of allocation, method of randomization, etc.), however important, would have led to a bottom effect compromising the usability of the instrument for differentiating between abstracts in the real world of a GI conference.

4. Points awarded based on design
   Points awarded based on design serve to counterbalance inequalities arising from the applicability of the various items. E.g. for randomized parallel controlled trials, to get a full score (2 points) for control for confounding, beside randomization, a statement on the comparability of relevant prognostic factors between the groups, or else, additional measures of control are needed. In
contrast, in cross over trials, comparability of confounding factors is assumed to be present, and 2 points will always be awarded. Consequently, if weighted equally, cross over trials would tend to score better than parallel trials. We appreciate that this is somewhat confusing, and I must admit that I have had problems putting this into words. Consequently, a section on this was dropped after the linguistic make over. I have now added a comment on this in the section "How to calculate a summary score".

I thought I had changed the reference to Dr. Cho, there seems to have been a mixup a different versions. My apologies.