Reviewer’s report

Title: Reviewer Agreement Trends from Four Years of Electronic Abstract Submissions

Version: Date: 11 October 2005

Reviewer: Liz Wager

Reviewer’s report:

General

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

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

It would be nice to have a more focused account about what is already known on the subject of reviewer agreement in the abstract. I felt this (and the Methods, 2nd para, p8) contained too much information about electronic submission, which was not really the focus of this study. Although electronic submission made the data collection feasible, it didn't seem fundamental to the study, so I felt the discussion about its merits was not particularly helpful (1st full para on p6). I would have preferred to see more detail about the review criteria, which are hard to judge from the short titles (eg Methods I and Methods II).

Given that BMC is an e-journal, I'd encourage the authors to append a review form (e.g. with criteria for RCTs -- I realise that you don't want to present all the different variants) as an appendix. This would also help readers judge for themselves which criteria are more or less objective.

I would like to see more discussion about what happened in 2003 and why the authors think this seemed to change both the mean scores and the degree of agreement. Do they think this was mainly due to the change in criteria? If so, why was the agreement not maintained in 2004.

The authors should discuss the possibility that the effect was either a random variation, or a non-specific effect of change. Any change in the criteria might make reviewers read them more carefully and therefore be more consistent, and this could explain why the effect 'wore off' the following year. For this theory to be true we'd also need to understand how many reviewers had reviewed in previous years and how many were 'new'. There is good evidence that any change can have effects (eg changing light levels alters factory production, but it doesn't matter whether the change is an increase or a decrease, it's simply the change that matters) -- this needs to be discussed.

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

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The abstract gives the impression that all abstracts were rated by just 3 reviewers (1st line of methods) -- I suggest this should be rephrased to read 'and each assessed by three (out of **) volunteer CAEP reviewers'.

The total number of reviewers should also be mentioned in the full text (somewhere in the Results section).
Overall, the MS is well presented, but I spotted a few typo's:
Ref 9: should be Intern Med (not Inern)
Ref 16: should be judgement (not judement)

A phrase on p.13 confused me. You state 'the total score EXCEEDS the 0.40 value' whereas I think you mean it crosses (but goes below) this.

Discretionary Revisions (which the author can choose to ignore)

One advantage of a database over several years is that presumably you can build up data about individual reviewer characteristics. Although this is mentioned in the discussion as a topic for further research, I would very much have liked to have seen something about the effect of individual reviewer characteristics on review scores. The definition of the types of variation (abstract / reviewer / abstract-reviewer, bottom of p9) is nice and clear (although maybe it would fit better in the Introduction than the Methods), but leads the reader to expect something about reviewer variation. Is it possible to know, for instance, whether some reviewers scored consistently harsher or more leniently than others (ie were there 'assassins' and 'zealots')?

I would have liked more information / discussion about the 6-point item for 'overall' impressions. At first sight this seems very subjective, yet it got quite good agreement. If that's the case, one might argue that getting reviewers simply to use a 6-point subjective score would be just as good as the 'objective' assessments. It would therefore be interesting to know the extent to which the score for 'overall' correlated with the total score for each abstract. This would also help to validate the scoring system.

What next?: Accept after minor essential revisions

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I declare that I have no competing interests