Reviewer's report

Title: Efficient clinical evaluation of guideline quality: development and testing of a new tool

Version: 2

Date: 10 February 2014

Reviewer: Joy C MacDermid

Reviewer's report:

In general this is a clear paper that addresses evaluation of clinical practice guidelines. This is an important topic given the large public investment that goes into development of clinical practice guidelines and the challenges that front line users often have in implementing them. Overall, the work has value and is worth sharing.

The authors of this work have developed a new tool for evaluating clinical practice guidelines which they expect to be equally valid to the AGREE II - which would I expect to be the most commonly used instrument ... or current standard. They provide a legitimate rationale for the difficulties encountered in using this tool being complex in the number of items and the scoring. However, it would be fair for them to acknowledge the challenges in developing a new tool when an existing one is so widely used an internationally recognized. Is it their expectation that guideline developers would be willing to switch to their instrument? Or is this something developed for local use that will inevitably be used locally, but not internationally? I know that these things cannot be accurately predicted but the issues around acceptance of a new measure should be discussed briefly, when talking about potential uptake of this measure.

I find it interesting that one of the major rationales for developing this new tool is that it is simpler. Why did they not ask participants which one was simpler, record how long it took them to complete each evaluation using the tool etc.- If feasibility and preference are important issues, assuming the tool is equally valid, they should have been studied.

There is not a large sample of guidelines or raters. This may introduce some imprecision. Further, it appears that higher-quality guidelines were used. The range of scores on the new tool is from 75-100%. In terms of instrument development it would be better to have a wider range of quality of guidelines use as test guidelines. This is further indicated by the fact that all the guidelines were recommended by the respondents. There were no confidence intervals included around the correlation coefficients, therefore the issue of precision is not clear. The authors need to address whether they had an adequate range of guideline quality, and a sufficient sample size to provide a valid and precise answer to their question.

In looking at the data the range of scores is larger with the AGREE. This should
be discussed. Does it potentially mean that it is more discriminative?. Table 4 has no confidence intervals. I find the figures hard to follow. In figure 2, there is a plethora of lines showing how the different tools did on the AGREE, but this adds little to the fundamental question about the validity of the new tool and is very distracting graph to look at. In the next figure (3), the labeling is very confusing it says the number of agreements but looks like percentage. Further, the data points are connected lines when if I understand this graph correctly these are isolated estimates and hence should be bar-graph’s. In general, all of the figures the clearer legends and a description of what the table or figure is presenting. I think there are too many tables, some are not informative and could be removed. For example, a table showing that everyone agreed that the guideline was acceptable, is not that informative-it could easily be stated in text.

Minor revisions
I believe there is a spelling error in the first line of your tool full text NOT fill text. It would be important to portray the tool in the way that you want others to use it.

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.

Declaration of competing interests: no