Reviewer’s report

Title: A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation

Version: 1 Date: 14 April 2014

Reviewer: Nancy berkman

Reviewer’s report:

Thank you for providing me with the opportunity to review this research concerning a new tool to aid reviewers in conducting GRADE assessments of the quality of evidence included to answer the research questions included in systematic reviews. I found the MS to be clearly written and easy to follow.

I have a number of concerns with the study.

1. The evaluation of the checklist does not include a comparison group. Therefore, we do not know to what extent the use of the checklist would improve the reliability of the assessment beyond what would occur using the background and training materials provided by the GRADE working group. The authors need to comment on this omission.

2. The report is not clear concerning how many reviewers answered each of the questions in the tool. I also would like more information on the background and characteristics of participants. Individuals have apparently been conducting systematic reviews for a number of years (2-10) and while they may not have used GRADE, we do not know their level of experience and instruction in grappling with similar issues based on other grading systems.

3. The tool does not include what the grader is to do with their answers to the checklist in relation to developing a domain score or the final grade. The user of the tool would want to know whether there is a relationship between agreement through the tool and the domains and final assessment.

4. Sometimes the tool appears to be referring to individual studies. Because this tool is intended to be helpful in evaluating systematic review findings, it should always refer to aggregated evidence. If it is asking the grader to review the individual studies then the checklist should consider the variation between studies.

5. If the agreement on an item is fair or slight, why is it retained in the instrument? The instrument is designed to be of assistance in evaluating quality of evidence through GRADEPRO. Wouldn’t retaining these questions have the potential of harm, rather than benefit?

6. My experience is that the assessment of bodies of evidence based on MAs of RCTs are the simplest and most straightforward of SR quality assessments. Of greater need are tools that can be of assistance in making qualitative assessments of trials or assessments of bodies of evidence that include RCTs and observational studies. This tool was not designed to be used for those goals.