Reviewer's report

**Title:** Impact of quality of evidence on the strength of recommendations: an empirical study

**Version:** 2 **Date:** 2 June 2009

**Reviewer:** Rafael Perera

Reviewer's report:

Minor Essential Revisions

- Model Choice and Description

The description of the models used in the methods need to be made clearer. At the moment it appears to read as if the manuscript uses 3 types of models for analyzing the association between evidence level (very low, low, and moderate) and strength in the recommendations (uncertain, weak, strong):

a) Generalized linear mixed model with a cumulative logistic link function and using panel member as indexing factor

b) Generalized linear mixed model with a cumulative logistic link function with no indexing factor (treating all recommendations as independent)

c) Proportional Odds model (treating all recommendations as independent).

Models (a) and (b) are in effect ways of performing an ordered logistic regression that is based on the assumption of proportional odds being valid. This needs clarifying as well as the final choice (a or b?) and the reasons for this choice. Simplicity appears to be a good reason for choosing b if no major information is gained by estimating the correlation structure within rater.

Although the discussion about the assumption of parallel regression functions for a POM model to work is adequate, there is no mention about the goodness-of-fit of the model or if any diagnostic plots were carried out to validate the final model. This is important as the main conclusions of the paper are based on the model being an adequate representation of reality.

In particular, the presentation of an “exploratory analysis” treating quality ratings as continuous might not be adequate. Figure 1 could incorporate a second Y axis with the Observed probabilities that would show that the model underestimates for weak conclusion and overestimates for uncertain and strong conclusion in the “low quality” group. This should at least be included as part of the discussion.

- Presentation of Data

It would be useful to have information regarding the agreement between raters on the overall quality given by rater (which the supplementary figures provide) and on that given by the Systematic Reviewer.
The Figures could also provide more complete information not on the marginals – number (and %) of votes on strength separately to evidence quality, but also on how these values are associated with each other as they have done on Table 1. This could be as simple as adding the number of moderate/very low/low for each strength graph and vice versa.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests