Reviewer's report

Title: Determining the interviewer effect on CQ Index outcomes: a multilevel approach

Version: 3 Date: 30 April 2010

Reviewer: Howard Degenholtz

Reviewer's report:

The authors have responded in detail to the previous review, however some questions remain.

Major Compulsory:

1. The abstract lists the number of dimensions as 15, but only 5 of 10 had significant higher level variation, but “Adjusting for interviewer characteristics did not effect the overall star assignment…regarding 7 of 12 quality dimensions”. This is confusing and inconsistent.

2. The authors have provided a table with ICC values, however this does not really address the fundamental question, which is how much variation can be attributed to the interviewer versus the facility. Ideally, the authors should present the variation (random effect) at each level, and the percent explained by the inclusion of explanatory variables.

3. Related to #2, the authors do not present the appropriate null model: ie a model with no random effects.

4. Ignoring the clustering by resident, my understanding is that a cross-classified model (interviewer*facility) should have two level 2 random effects. The results present only one set of ICC values. Thus, it is impossible to interpret the findings regarding the magnitude of variation attributable to interviewer effect as opposed to facility clustering. Incorporating the resident clustering would add another random effect. It would help to provide an equation or set of equations for the different effects. This could be an online appendix.

5. The incorporation of institution level effects is not described on p. 8.

6. The definition of ‘experienced interviewers’ is flawed. The authors consider only experience within the present study, not prior experience. Thus, a new hire with no previous healthcare interviewing or exposure to working with the elderly is considered equivalent to one with 5 or 10 years such history. Second, the study does not examine whether experience itself has an effect, just whether there is an interviewer effect AMONG experienced interviewers. This should be clarified. The term experienced should be dropped for something more accurate, that conveys that these individuals had a high volume of interviews.

7. The description of the fieldwork is still not sufficiently detailed. On p. 7, last sentence of the first paragraph, authors write ‘…they performed their interviews in different combinations.’ How many times was each resident interviewed? Did
1. The order of interviewers vary? How many facilities did each interviewer work in? How much time elapsed between interviews?

8. The long discussion of the assignment of stars (pp 9-10) is mostly irrelevant: these findings are not presented, nor are they truly relevant for this purpose. The bottom line finding is that the ranking does not change. If that were not the case, it would make sense to demonstrate graphically the changes.

9. The -2loglikelihood presentation on Table 5 is not clear. What is being compared? Are any comparisons significant? How many degrees of freedom are there?

10. The table 5 results do not show all interviewer variables for each quality dimension. Why is this? Did the authors screen variables?

11. It is noteworthy that interviewer characteristics were only significant in 2 or 3 of the models; this could easily be the result of chance alone given the large number of comparisons.

12. The evidence of clustering within interviewer is modest, and more likely suggests that certain scales are open to more interpretation than others.

13. There is no theory provided to predict the presence or direction of interviewer effects. Thus, the finding that only a very few models have a significant effect suggests this is noise and not a major concern. The comments on the bottom of p. 13 go beyond the evidence presented.

14. The finding of minor changes in ranking is overstated on p. 14. While it may be important to the facilities affected, it does not seem to be a significant policy issue.

15. The suggestion that more than one interviewer conduct interviews in each facility will obviously increase cost; this should be discussed.

16. The last sentence in the 3rd para on p. 14 regarding absolute minimum interviews is not supported by these data.

17. The limitations section is not robust. Further details regarding the quality of self-reported interviewer characteristics, poor reliability of several of the quality dimensions, and lack of variability in others should be noted. Address the potential for the pattern of findings to be the result of multiple comparisons.

Minor:

1. P. 3, do not use parenthesis to set off the words ‘family of instruments’. Just state what it is.

2. P. 6, top line, do not being a sentence with a numeral.

3. On Table 5, please provide a note explaining the ‘type of care’ variable. What is the reference category?

4. The lengthy table 5 could be summarized in one large table with indications for whether the set of variables was statistically significant, rather than each variable. Detailed tables can be an appendix.

5. The finding on Table 6 could be combined with Table 5. Separate tables could be presented for random effects and fixed effects.
Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

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