Reviewer's report

Title: KIDMAP Developed on Web for Gathering Patients' Feedback on Their Doctors

Version: 1 Date: 13 January 2009

Reviewer: Andrew M Garratt

Reviewer's report:

Major Compulsory Revisions

1. Introduction. The authors state that the methods for assessing patient views at the physician level are not as well established as those at the organisation level. This may well be the case but some further background is necessary as well as evidence to support this assertion. What exactly is the difference between these two approaches? Some examples to illustrate might help here. There are a number of questionnaires available that measure patient views of their physician. Moreover, the quality of the interaction (communication etc) with the doctor is a key component of questionnaires that have been used to assess the quality of care delivery at the organisational level.

2. Introduction. Assessing doctors to be increasing prominence. The third sentence of this paragraph is rather unclear. What exactly is a hospital initiative? I would argue the contrary to what I think is being stated here. That is, hospitals rarely assess patient satisfaction as part of routine management. Such work is usually research based or takes place infrequently as a component of national or regional quality improvement initiatives as in England and Wales with Picker and the US with CAHPS.

3. Page 4, para 2. CTT does not require that all items are completed in order to produce a final score. Imputation is used for missing data. What is more relevant in the context of a web-based survey is that "CTT requires that all items within a questionnaire are given to respondents in order to assess their satisfaction". This is by far the most important contribution that modern psychometrics including IRT and Rasch analysis can make in the context of electronic administration and what is referred to in the literature as Computer Adaptive Testing (CAT). What role does CAT have in the context of patient satisfaction measurement? This is worthy of some discussion. Was it not considered for use here? This could be a future development relating to the work reported. Such a method has obvious relevance in the context of health status and quality of life assessment but its value in the context of patient satisfaction is less clear. In relation to quality of life measurement, CAT has been used to determine which questions are most relevant to a patient based on their responses to preceding questions. Such a method may have relevance to the extent that patient satisfaction questions form a hierarchy starting with questions that patients are least likely to be satisfied with to those that they are most likely to be satisfied.
4. Picker 23-item questionnaire fitting to Rasch model. The first sentence is not clear. Was this questionnaire based on a review of ten existing questionnaires? How can the KIDMAP assess whether the patient is "fumbling, careless, or hoaxing to answer" questions? Is it not more accurate to say that the patient has not answered the question as predicted by the model based on the response of other patients? 

5. It is stated that the questionnaire has been tested. What were the results of reliability and validity testing including Cronbach's alpha and any comparisons with existing instruments or questions relating to satisfaction? This only requires a 2-3 sentences as a summary.

6. Table 1 shows that the 23-items often include more than one aspect of doctor performance. While these are related it is possible that a doctor performs well in relation to one aspect of the question but not another. For example, in relation to question 10 a patient may have received help understanding their illness but not their treatment. Splitting this question may be a better alternative.

7. Page 7, para 5. "more valuable" should be changed to "potentially more valuable" because this has not been tested and as such there is no evidence for the value or usefulness of such information to doctors or other groups for purposes of quality improvement initiatives.

8. Participants and procedure. What was the process for administration of the questionnaire? Were those that completed the questionnaire given help? Were patients administered the questionnaire in a consistent fashion? ‘Newer’ methods of questionnaire administration are given in the Introduction - computer, for example. How do these tie in with the present study? Further information is required here. Were the methods of administration piloted?

9. Page 10, para 1. There is a need for further information relating to the interpretation of results. I presume that "easy" means that the items showed that patients were more satisfied than dissatisfied according to the five-point Likert scale. This needs clarification to aid the reader.

10. Discussion relating to the implications of DIF in relation to age and disease groups. There are a large number of items within the scale already so why not consider removing those that exhibit DIF.

11. The interpretation of KIDMAP requires training of doctors. In the Discussion it is stated that the display of a diagram can also help reduce patient burden. Such a system based on the results of IRT is not going to be understandable by patients without some education package alongside. In the Discussion it is stated that a child can understand KIDMAP. Further explanation is needed. Some of the text might be better placed in the Methods section.

12. What is the doctor's self-rating?

13. The English could be greatly improved including the general structure, use of
subheadings, sentence construction and selection of words. It is recommended that the authors use one of the many services now available for this task.

Minor Essential Revisions

1. "Accessible" (page 4, first para) requires definition. Accessible to whom, easier to complete, user friendly, etc? I am not sure how the statement that patients are deriving considerable benefits from using the internet follows from the use of an internet based survey.

2. Introduction, para 4. The second sentence applies only at a general level and a possible amendment would be as follows: "Questionnaires are usually distributed either via a consecutive...and respondents usually return a questionnaire..."

3. "Assesses" should be used instead of "reviews" in the Abstract Objective. The second sentence in the same paragraph is difficult to follow.

4. The start of the first sentence of the Results section of the Abstract needs reworking eg "21 of the 23 items fitted the Rasch model..."

5. Page 4, para 2. The example of mental health is not necessary when patient satisfaction is used which is also the context.

6. Page 4, para 3. "Outliers" is recommended in preference to "Abnormalities".

7. Page 6, para 1. "Skilfully" should be replaced with "The questions were scaled using a five-point Likert scale..."

8. Page 6, para 2. What is meant by diagnostic?

9. It is important for readability that the authors are consistent with terminology throughout the article. Test theory terminology is sometimes used that is unnecessary and potentially confusing eg "examinee".

10. Page 6, para 5. "Distinct points of view" is not a good phrase in this context. "...groups respond to individual questions differently irrespective of their satisfaction" is perhaps better.

11. Discussion. I would prefer "visual representation" to "diagnostic diagram" within this context.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I declare that I have no competing interests.