Reviewer's report

Title: Online Assessment of Patient Views on Hospital Performances Using Multivariate instead of Univariate Approach

Version: 1 Date: 7 April 2009

Reviewer: Moritz Heene

Reviewer's report:

Major Compulsory Revisions:

The paper addresses a practically important question of patients evaluation of healthcare. Moreover, the study overcomes the methodological ill-founded data analyses and interpretations of previous studies. Its style is concise and to the point, well-argued and clearly set-out. However, there are some methodological problems which need to be discussed:

Paragraph “Data collection”:

“The original data for the 88,308 patients were not available to us. Instead, the aggregated data for the 169 hospitals were analyzed.”

- It is completely unclear how an item response model –which is by definition based on original response data- can be applied to aggregated data. Furthermore, it is not explained how the authors aggregated the data. This paragraph needs a thorough revision.

Paragraph “Data transformation and analysis”:

- As mentioned by the authors Andrich’s rating scale model was applied. This model strictly requires ordered category threshold parameters. In the case of reversed thresholds the data does not fit to the model. I would therefore recommend to report whether reversed thresholds did occur.

Paragraph “Results”:

- Please explain the general principle of the principal component analysis of the Rasch residuals. For readers not familiar with this method it is unclear what these results actually mean and whether they add anything new in comparison to the item fit statistics.

- Item fit statistics: The authors only report results from the infit mean square statistic. Please also report and discuss the results from the outfit mean square statistic which covers important information about so-called off-target, i.e. unexpected, responses. While the software they use for their study provides the outfit statistic by default, the results produced for fit diagnostic is not presented in the paper and limits its significance.

Minor Essential Revisions:
Paragraph “Methods”: “We used the Rasch model (1960)…”
- Note that Rasch (1960) refers to the dichotomous Rasch model, i.e. an item response models for binary coded items. The authors used the rating scale model by Andrich (1978) which is a generalization of the Rasch model.

Paragraph “Inappropriate individual item analysis”: “…global relationship and interaction between items”.
- It is unclear what is meant by “interaction”. Usually, item interaction refers to unwanted aspects of item responses like violations of the local stochastic independence assumption. I would therefore suggest to omit the term “interaction”.

**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.