Reviewer's report

Title: Configural frequency analysis as a method of determining patients' preferred decision-making roles in dialysis

Version: 2 Date: 28 May 2010

Reviewer: Hugo J. Duivenvoorden

Reviewer's report:

Major compulsory revisions

On page 7 the authors present a number of advantages of CFA. They state that CFA “...requires only variables on a nominal level ...”. I doubt whether this is an advantage.

You could formulate in terms of ...CFA is suitable for nominal data.

Furthermore, the authors state that: Another problem with typal analyses based on correlations (like factor analysis) is that they take into consideration only first-order interactions of the variables”. This statement seems confusing. For example, factor analysis is usually performed on zero-order, bivariate correlations.

On page 9 is reported that dialysis units (=centres?) received 14972 questionnaires. As the individual centres did not keep count of how many survey questionnaires had actually distributed to their patients, the response rate could not be calculated. Of the questionnaires distributed 6318 were returned for statistical analysis.

But, it might be of relevance to present the “response rates” for each individual centres, based upon the number of questionnaires the 187 dialysis centres received. Do the “response rates” differ for the different centres? What is the median value (interquartile range, minimum, maximum) of these “response rates”?

Possible effects of different centres have to be discussed.

On pages 10-12, the (semi-quantitative) measures were discretized into dichotomisations or trichotomisations. The question arises whether discretising variables induced loss of information. And why they have dichotomized three variables and two variables trichotomised? Furthermore, I don’t understand why the authors have used different ways of discretizing: sometimes the mean, sometimes median, sometimes otherwise. It would be helpful if the authors describes the rationale of different ways of discretising the variables.

The (semi-quantitative) measures are clearly non-nominal, while the authors state that one of the advantages of CFA is that the variables are on a nominal level (see Page 7).
On page 14, “Sample characteristics” the authors try to show that the study sample is representative of the overall population. The study sample was compared with the overall population with regard to gender, age, educational level, length of dialysis and Karnofsky index. The question is whether these characteristics are of real relevance. It could be that the relevant factors are psychological of nature. The motives to participate or not might be of great importance, which may introduce selection bias. I think that this aspect should be elaborated in the “Discussion”.

Minor essential revisions

On page 9 the authors basically describe the study design. As 187 dialysis centres participate to contact the patients, the type of design is multilevel (level 1: patients; level 2: dialysis centres). In the manuscript the effect of level 2 was not taken into account. I would appreciate if the authors discuss this aspect of the design and possible threatening of valid statistical inferences.

On page 10 five predictor variables are mentioned, one of them is age. Age as predictor variable was not discussed in the introduction. I would suggest to describe the rationale that age was considered as predictor variable.

On pages 12-13 CFA is described. As most readers are not statisticians, I would appreciate if the authors give a clear description of CFA in only nontechnical terms (i.e. without statistical terms and without symbols), while a technical description, including a clear explanation) is presented in the “Appendix”. The technical part of the description on Pages 12-13 should be moved to the already presented “Appendix”.

For sake of good and easy understanding –once again, most readers are non statistician- I would suggest to describe the section “Analyses” in nontechnical terms (pages 12-13).

The phrase on Page 15 “There were also no differences observed between the exploratory sample and the confirmatory sample due to the random division of the sample” needs reformulation. I would suggest something like: “As the patients were randomly assigned to either exploratory or confirmatory sample and as the two subsamples were large, it was expected that only differences due to random error would be found. These results were in line with this expectation.”

In Table 2, on Page 31, I believe that the codes j1 to j6 don’t have any additional value. Verbal explanation of the codes of the six variables is sufficient.
In addition, the authors may consider to present the real percentages (instead of the p-values or add to the p-values).

On page 18 the algebraic notation of the predictor-criterion configurations do not “read” easily. I prefer a statement in nontechnical terms.

Regarding the “Discussion” section, the length of it may be shortened.

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

**Quality of written English:** Acceptable

**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