Reviewer's report

**Title:** Psychometric properties of the Iranian version of 12-item Short Form Health Survey (SF-12): A population-based study

**Version:** 1  **Date:** 25 June 2009

**Reviewer:** Ulf Jakobsson

**Reviewer's report:**

An interesting paper with focus on psychometric evaluation of the SF-12 instrument. The study have a sufficient number of respondent included in the study and the sample seems fully adequately chosen. However, the authors only perform some of the possible analyses that should be conducted to fully evaluate the instruments psychometric properties. For example, to only use Cronbach’s alpha to assess reliability do not give any firm knowledge about the instruments reliability. Furthermore, the authors have only evaluated construct validity in the study and all other types of validity measures (such as face validity, content validity, criterion validity) is not considered/discussed at all. I suggest that e.g. corrected item-total correlations, analyses of discriminant and convergent validity, analyses of known-group validity between other groups than only age and gender are added (and maybe also add a discussion about face & content validity).

**Major Compulsory Revisions**

(1) The background is too sparsely written and do not have any argumentation for the need of this study. Furthermore, there is no critical discussion about the results from previous studies, for example the authors state that the instrument is a valid and reliable measure, but this is can be questioned. All results from previous studies do not indicate this.

(2) In the method section, the instrument is too sparsely described. More information about e.g. the scoring method and development is needed. Furthermore, the paragraphs under the heading “The questionnaire” and “Translation” should be merged.

(3) Why an explorative factor analysis instead of a confirmatory factor analysis is used is unclear (the factor structure of the SF-12 is well-known). The authors need to more explicit explain and discuss the reason to choose an EFA in favor to an CFA. Furthermore, more information is needed about how the EFA was done (including extraction method and how various “goodness-of-fit tests” was applied). To be noticed, the measurement model (both SF-12 & SF-36) is somewhat complex with cross-loadings which is hard to identify with an EFA.

(4) Table 3. Figures regarding eigenvalues, communalities, and explained variance are missing.
(5) The use of Pearson's correlation does not seem appropriate to use when correlating various items with the summary scores since the items are nominal and ordinal data. A corrected item-total correlation should be used instead (based on appropriate correlation matrix).

(6) The discussion is in general too weak in the discussion of the results as well as the methodology and study limitations. It should also more thoroughly discuss the limitations related to the chosen methods for psychometric evaluation. Why only these methods used and what are the consequences of this.

(7) The conclusions (both in the discussion section and in the conclusion) about that the instrument is found to be valid and reliable needs to be revised because the results from the study do not fully support the statements made by the authors (e.g. the floor & ceilings affect is not acceptable).

Minor Essential revisions

(8) The analyses of floor and ceiling effect is not mentioned in the method section and should be inserted under the subheading Statistical analysis.

(9) In the discussion section, the authors state that there is no problem with floor and ceiling effect. However, this is not supported by the results.

(10) Table 2 & 4 & 5, the mean values should not be used for the subscales (due to several subscales is only based on one item and all items in SF-12 is either nominal or ordinal data).

(11) Table 4 & 5, the Student t-test and ANOVA should not be used for comparisons regarding the subscales (due to several subscales is only based on one item and all items in SF-12 is either nominal or ordinal data).

(12) I think that the authors should be a little cautious with the statements about that the results could serve as normative data (for the entire Iranian population).

Discretionary Revisions

(13) Based on that the psychometric evaluation of the instrument is limited to only construct validity, the authors may consider changing the title in accordance to this. If the authors chose to not change the title, then consider removing “a population-based study” in the title since it do not add any important information.

(14) Readers will benefit from a revision of the layout in the tables (so they will be easier to read, all necessary information is included, and the data is presented properly).

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

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.