Reviewer’s report

Title: Psychometric properties of the Korean SF-12 v2 for assessing the general population

Version: 2 Date: 9 September 2014

Reviewer: Sepideh Omidvari

Reviewer's report:

Major Compulsory Revisions:
1. Methods, Paragraph 4 (Measurements): what do the authors mean by “the official Korean SF-12 v2”? It needs more explanations. And what about “face validity” of Korean version of SF-12 v2?

2. Results, Paragraph 4 & Discussion, Paragraph 3: The internal reliability of the MH scale was low (0.37) in the study. How can it be explained?

3. Results, Paragraph 4 & Discussion, Paragraph 3: It has been mentioned: “Item factor analysis revealed the presence of three factors that accounted for 65.1% of the variance. The PF, BP, and GH items loaded onto the physical health concept (factor 1) and the VT, MH, and GH items separately loaded onto the psychological health concept (factor 3). The SF, RP, and RE items loaded onto factor 2” and “This pattern is unique to Korea and seems to reveal a different view in Korean residents, as the RE and RP items were also loaded onto the same factor in the SF-36 v2 in Korea”.

How can the results be explained? What do the authors mean by “different view in Korean residents”?

4. Discussion Paragraph 1: It has been mentioned: “all items were more highly correlated with other competing components than their own hypothesized components”.

It seems a mistake and the sentence is not consistent with the results.

5. Discussion, Paragraph 3: How can “higher ceiling effects” be explained in this study?

6. Considering “Background, Paragraph 2” (Our current study aim was therefore to evaluate the psychometric properties of the Korean version of the SF-12 v2 using a general representative population) and “Methods, Paragraph 2” (The target population was individuals aged 19 years or older living in Korea (except Jeju Island) who consented to the survey participation. Sampling was performed using a multistage stratified quota method. Sample quota were assigned to each of 15 Korean regions according to population structure (gender, 10-year age group, and level of education [12 years or less vs. more than 12 years]) as defined by the resident registration data of the Ministry of Administration and Security of South Korea in June 2013), how can paragraph 4 of discussion section be interpreted (There were some limitations to our present study. In the
first instance, even if we had recruited respondents nationwide, the external validity of the sample would be limited. The age and sex distributions of our sample were similar to those reported in the 2010 national census but participants in this study reported lower health care utilization than the participants of the 5th KNHANES, which is a national-wide health survey of more than 30,000 people. This may impact the HRQoL by producing high item scores and a low floor effect?

7. Some parts of the paper are similar to the article “Kim SH, Jo M-W, Lee S: Psychometric Properties of the Korean Short Form-36 Health Survey Version 2 for Assessing the General Population. Asian Nurs Res. 2013, 7:61-6”. The paper needs to be revised.

Minor Essential Revisions:
1. It seems that the title of paper needs some changes (the phrase “assessing the general population”).
2. Background, Paragraph 1: It seems that it is better to change the word “recently” to “recent decades”.
3. Background, Paragraph 1: The word “comprehensive” may not be appropriate to describe SF-12.
4. Abstract, Paragraph 2: “testing the feasibility and understanding of the psychometric properties of the instruments should precede their use in research when instruments developed in other countries are adapted to the Korean circumstance” has been mentioned in the paragraph. What about the feasibility of SF-12 in the study?
5. Methods, Paragraph 2 & Discussion, Paragraph 4: Considering the sentence “we did not explore concurrent validity, test-retest reliability, and responsiveness”, what do the authors mean by “Out of 3,206 households that were contacted for interviews, 1,000 successful interviews were conducted (31.2%)”?
6. Methods, Paragraph 4 (Measurements): It has been mentioned: “The 12 items are used to derive two summary measures (i.e. physical component summary [PCS] and mental component summary [MCS]) and a preference-based health utility index”. What about the “preference-based health utility index” in the paper?
7. Results, Paragraph 3: It has been mentioned: “Significant differences were observed in SF-12 v2 scale scores. As expected, the scale scores of women were significantly lower than those of men in all scales except the SF and RE scales. The oldest age group (# 60 years) demonstrated a significantly lower value than the other age groups on most scales except the MH scale when a post hoc Tukey comparison was applied. Higher educated people tended to report higher values than poorer educated people on all scales. People suffering from disease and who recently used the hospital service demonstrated significantly lower scores than other participants on most scales.”

What about the P values?
8. Discussion, Paragraph 1: “Generally, the item scores in our sample were
higher than in other countries”
How can it be interpreted?

9. Abstract, Paragraph 3: The “Results” part is lacking information about "reliability" of the measurement.

10. Abstract, Paragraph 4 & Conclusions, Paragraph 1: What do the authors mean by “practical” (...the Korean SF-12 v2 is a practical, valid, and reliable instrument)?

11. What bout marital status of the participants?

12. Table 2: What does “NA” (in the first line) stand for?

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

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

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.