Author's response to reviews

Title: Assessment of psychometric properties of the Korean SF-12 v2 in the general population

Authors:

Seon-Ha Kim Dr (kshgive@gmail.com)
Min-Woo Jo Dr (mdjominwoo@gmail.com)
Jeonghoon Ahn Dr (jahn@neca.re.kr)
Minsu Ock Mr (ohohoms@naver.com)
Sangjin Shin Dr (jshin@neca.re.kr)
Jooyeon Park Ms (jypark@neca.re.kr)

Version: 3
Date: 6 October 2014

Author's response to reviews: see over
Dear

We appreciate the Editor’s & Reviewer’s time as well as all of the comments, which contribute significantly to the revision of our manuscript. I attached my answer for each comment. Our manuscript has been edited by professional English editing agent. We have highlighted red changes in the manuscript.

---

**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?

   - Official version means Korean SF-12 offered by organization owned copyright as it is. We deleted ‘official’ to prevent confusion.
   - We did not perform translation process.
   - Face validity was important component as your comment. The SF-36 v2 have formally been translated by Quality metrics ©, thus, our study rather focus on data quality evaluation of SF-12 according to developer guide in Korean context, however we did not explore face validity of the SF-12. We added this fact the limitation of the manuscript.

   Page 8 Line 13-14: we did not explore face validity, concurrent validity, test-retest reliability, and responsiveness for health state change

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?

   - We appreciate the Reviewer’s comments. Korean people may think differently in two MH items (Calm & peaceful, downhearted & depressed) resulting in low internal consistency. In Korean SF-36 study, these two items were loaded to different factors [9]. In Chinese study, showed similar result (alpha=0.34) [23]. In Iranian study, Cronbach’s alpha in MH (0.61) lower compared with other scales [Rohani et al, Iranian rehabilitation Journal, 2010: 8(12)] Participants may respond mistakenly. Other scale items showed same direction, which is higher level indicated better condition. However MH scale items have opposite direction among two items.

   - We added one sentence in discussion section of the manuscript as below.
   - Page 7 Line 29-31: Korean people may be free from the influence of two MH items (Calm and peaceful, downhearted and depressed), respectively.

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”?

   - We appreciate the Reviewer’s comments. “different view in Korean residents” sentence
were deleted due to unclear meaning. RP in this study was not included in psychological factor, this is different from western people. SF, RP, RE could consider the third factor neither PCS nor MCS. Factor

We added a sentence in discussion section of the manuscript as below.

Page8 Line 45: Use of item or scale scores rather than two summary measures of the SF-12 v2 seems to be more appropriate in Korea

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.

We appreciate the Reviewer’s comments. It was my mistake. I revised “all items were more highly correlated with their own hypothesized components than other competing components” in line 31-32 of page6.

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

In table 2, percentage of highest level (level 3 in PF items and level 5 in other items) could explain ceiling effect. Higher level score indicates to better condition. Compared to study in other countries, percentage of the highest level in our study tended to be higher than other studies.

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)?

We appreciate the Reviewer’s comments. We deleted the word ‘representative’. As you know, quota sampling have limitation compared to random sampling. We considered sample quota (age, sex, and level of education, however, we could not consider all components (such as health utilization). That is the reason we describe limitation and its potential effect.

We added a sentence in discussion section of the manuscript as below.
In line 10-12 of page 8, Lower health care utilization may indicate that our sample was healthier than Korean population. Healthy people may assign the HRQoL score ~

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.

We appreciate the Reviewer’s comments. I know that some part of Kim et al’s article are similar to my current article. Kim et al (2013) was my article using similar research method and showed similar results. The only different part was HRQoL instrument (SF-12 vs. SF-36),
so inevitably, two articles have similar parts. I revised some words or cited reference [9] in the manuscript. But still somewhat similar part exists. I already checked it using self-plagiarism program before submitting this article. Agreement rate of revised manuscript decreased from 18% to 12%. And I think that the remained similar part was not problematic.

**Minor Essential Revisions:**
1. It seems that the title of paper needs some changes (the phrase “assessing the general population”).

   I change my title to “Assessment of psychometric properties of the Korean SF-12 v2 in the general population.”

2. Background, Paragraph 1: It seems that it is better to change the word “recently” to “recent decades”.

   As your advice, I changed the word “recently” to “recent decades”.

3. Background, Paragraph 1: The word “comprehensive” may not be appropriate to describe SF-12.

   I revised the phrase “comprehensive measure” to “measures various aspects of health status” in line 8 of page3.

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?

   The feasibility of SF-12 in the study was measured using completeness of data and out of range data. The completeness of the data was 100% and there were no out-of-range values. These already described in method and result section.

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%)”?

   I guess you misunderstood ‘responsiveness’ as ‘response rate’. Responsiveness has been defined as the ability to detect changes that are meaningful or clinically important. [Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care. 2000, 38:84-90]

   In order to clarify the meaning I added ‘for health state change’ to the sentence in line 14 of page8.

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?

   The part was explanation of SF-12. SF-12 could generate a preference-based health utility. However, our study did not generate preference-based health utility because Korean tariff of SF-6D did not currently exist. In order to prevent reader’s confusion, I deleted ‘preference-based health utility’ in measurement section of the manuscript.
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?

‘(P>0.05)’ was added in Table 4 Footnote. I only marked non-significant difference, because, most comparison in t-test or ANOVA showed P <0.05.

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

It could be Korean people evaluated positively regarding their own health than other countries’ people. I added a sentence ‘Korean people seem to evaluate themselves as healthy compared to people from other countries’ in line 1-2 of page7.

Or our sample could be healthier than subjects in other countries. That’s why I described the limitation of our sample.

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

“The reliability of all SF-12 v2 items was 0.88” was added in abstract. “The reliability of all SF-12 v2 items was 0.88. Cronbach’s alpha for the PF, RP, GH, and BP items was 0.83, and that for the MH, RE, VT, and SF items was 0.79” in line 15-17of page 6 was added in result section.

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)?

I changed ‘practical’ to ‘feasible’ in order to clarify the meaning.

11. What about marital status of the participants?

We added marital status information in Table 1.

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

NA means ‘not applicable’ I added abbreviation in Table 2 footnote. Two items have three response options and other items have five response options.