**Author's response to reviews**

**Title:** Health related quality of life measured by SF-36: a population-based study in Shanghai, China

**Authors:**

rui wang (fengxinzimail@hotmail.com)  
cheng wu (wucheng_wu@hotmail.com)  
yan fang zhao (zhyf115@hotmail.com)  
xiao yan yan (ghost_tian@hotmail.com)  
xiu qiang ma (mxqiang1823@hotmail.com)  
mei jing wu (wenny2700@hotmail.com)  
jia he (hejia63@yahoo.com)

**Version:** 3  **Date:** 12 June 2008

**Author's response to reviews:** see over
Dear Drs. Iratxe Puebla and Rikki Graham:

Thank you very much for your consideration on our work. We appreciate the valuable comments from the two reviewers. We have revised the manuscript accordingly. Attached please find our responses to the suggestions and comments.

I hope that you will find the revised manuscript acceptable for publication in BMC, and am looking forward to hearing from you soon.

Best wishes,
Rui Wang

**Reviewer:** Dr. Giovanni Apolone

**Major compulsory revisions**

1. As Authors have carried out this study with the primary purpose to test the feasibility and “actual” validity of this questionnaire in the real setting/world, it is quite disturbing that they did not perform any kind of analysis to evaluate the frequency and the pattern of the non-responder status. As a matter of fact the prevalence of non-responders was quite high (only 919 cases were actually used for the final analysis, yielding a 23% proportion of non-responders). This fact reinforce the need to describe the comparative difference between responders and non-responders and to identify the risk factors of the non-responder status. In addition, given the fact that it was possible to help/assist people, this event also should be quantified and evaluated as it is well known that people asking/receiving help are a quite peculiar sub-sample and help may distorce results. An analysis of the pattern of missing values at item level may also help identify critical/disturbing items.

**Response:**
We randomly sampled 1200 subjects and 1034 of them actually answered the questionnaire. The overall response rate was 86.2%. After the field work ended, for the purpose of quality control, we randomly called 10 respondents from both Huangpu and Songjiang survey sites, respectively, to see whether or not they actually received the interview. They did. But when we randomly called 10 respondents from Pudong district, we found that the respondents from the same residential area interviewed by an interviewer, Mr. Linxiang Huang, did not receive the interview actually. It turned out that Mr. Huang didn’t interview those subjects. After having discussed with the sponsor, principals of the survey site, we deleted the data for the 112 subjects interviewed by Mr. Linxiang Huang. In the survey process, we applied the district-block-residential area sampling method, and, then, randomly sampled the
subjects from each residential area. Each interviewer was in charge of only one residential area. Therefore, deleting the 112 respondents above would not affect the representation of the whole sample. In addition, three questionnaires were withdrawn from statistical analysis because more than 80% items were uncompleted.

Interviewers did explain the questionnaire to the respondents when they had any questions. But all the interviewers were trained by the same experts and every questionnaire item had a standard explanation without inducement. Hence, we believe that the interviewers' explanation had no influence on the results.

The missing values in the SF-36 dimensions were imputed as follows: if 50% or more items in one dimension were completed, the mean value of the completed items was used to impute the missing values. If more than 50% of the items were missing, the dimension score was excluded from the statistical analysis. In our survey, the item response rates were actually quite high. The average item response rate of the general information was 98.81%; the average response rate of the 36 items in SF-36 was 99.67%, ranging from 98.80% to 99.89%.

2. Results document that some scales had quite low reliability estimates, such as the SF. Although in other countries and languages this fact was already documented, it requires a deeper analysis and discussion as other findings, such as the low performance of the MH scale (one of the best performer in other settings), suggest that there might be some problems in the very conceptualization of mental health in this sample.

Response:
The low reliability of the SF dimension was found world-wide. MH and VT also had relatively low reliabilities. The Cronbach's $\alpha$ coefficients for the VT and MH dimensions were 0.66 and 0.75 in the survey of Hangzhou, 0.72 and 0.71 in Sichuan, 0.74 and 0.77 in Hong Kong, 0.73 and 0.74 in American Chinese, and 0.78 and 0.69 in our study, respectively. This may be due to the characteristics of Chinese people. They are not used to talking about their feelings and emotions in public.

3. Some multivariables analysis using as dependent variable the GH scale (that has a generic meaning and could be related to either a physical or mental concept of health) and as independent factors the other scales could improve the understanding (better than the factors analysis) of the complex relation between physical and mental scales and global health

Response:
GH (general health) may be related to either physical health or mental health, but it is correlated to physical health more closely. We think that the 8 SF-36 dimensions had the equal status and the correlation analysis demonstrates the relationships between each
dimension and PCS/MCS.

4. Somewhat disturbing is the fact that female in this sample reported almost always higher scores that males, across relevant subgroups. Despite the fact that multivariable analysis reduced the magnitude of this phenomenon, this fact should be carefully studied and discussed as in other settings it was observed the opposite relation.

Response:
We noticed that female always had lower scores than male in the SF-36 dimensions. In China, women have pressures from both family and society. They are asked to be a good wife, a good mother, and a good worker. The pressures may lead to a worse quality of life. Further investigation on this issue will be conducted in future studies.

5. The poor global performance of SF is also very disturbing. Although noted in other settings, in this context together with other findings it suggests that something of the mandarin version of the SF-36 is wrong, at least in this sample. The scale is formed by two very similar questions and usually the item-item correlation is high and problem comes from convergence-discriminative analyses. In this case the item-item correlation is quite substantial (0.70), but the reliability coefficients are low and apparently no problems are evident in terms of correlations with items from other dimensions (range 0.110-0.504). Authors suggest that an explanation of the poor reliability may be the reverse orders of the two items. It could be the case but the high item-item correlation does not support completely this explanation.

Response:
The SF dimension had the lowest Cronbach's $\alpha$ coefficient in this study, which was consistent with other surveys using the Mandarin version of SF-36. The SF dimension also had the lowest ICC and split-half reliability, indicating there might be some problems in the conceptualization of social function. Traditionally, Chinese people don't think much about social function, and commonly say little about how the physical health or emotional problems interfere with their social activities. In addition, the SF dimension included two questions as follows: (1) “during the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?” and (2) “during the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?” It appeared that the answers of the two questions had reverse orders, which may lead to the low reliability. The other reason may be the cultural diversity. In China, “social activities”, translated as “she hui huo dong”, refer to not only the everyday life within a family or one's circle of friends, but also the formal activities with other people such as going to a ballroom dancing event or attending a conference. The misunderstanding may result in the low reliability.
Minor essentials

1. It is not clear what kind of chronic diseases were evaluated. Only physcials? Do Authors have some information about the presence of mental/psychologic problems in the sample studied?

Response:
Respondents were asked whether or not they had been diagnosed by physicians with the following chronic conditions: hypertension, ischemic heart disease, cerebrovascular disorder, diabetes, chronic obstructive pulmonary disease (COPD), asthma, renal disorder, liver disorder, rheumatoid arthritis, osteoarthritis, anxiety, and depression, and at which age the disease had been first diagnosed. The use of medicine at the time of the interview was also recorded.

2. Did Authors use the first or second release/version of the SF-36 HS? Did they use the RAND or IQOLA rules and algorithms?

Response:
We used IQOLA SF-36 Standard UK Version 1.0.

3. Scores from scales are quite high, when compared to other studies. Authors should discuss these findings in a deeper way, as it could be another indicator of a difference between Chinese and non-Chinese “culture” and conceptualization of health.

Response:
Compared our survey with other studies, it was evident that American Chinese had the worst quality of life among different Chinese populations. Shanghai population had the best quality of life, even better than American and Canadian. It should be noticed that the other studies in the comparison were undertook much earlier than our survey, and China has made impressive progress in living standard during recent years. Especially, Shanghai is the financial and commercial center of China with the best medical and sanitation conditions. For example, the average life expectancy of the Shanghai population was 81.08 years old in 2007, which is slightly lower than the average life expectancy of Andorra, Macau, Japan, Singapore, San Marino, Hong Kong, and Canada. The infant mortality rate was 3.0‰ and maternal mortality rate was 6.68 deaths per 100,000 live births. All these factors may lead to high HRQL in the Shanghai population.

4. Reporting and analyzing the summary scores (PCS and MCS) estimates overall and across groups may help understand some of the phenomena above identified
Response:
We have added the analysis on the summary scores (PCS and MCS).

5. Authors should add in the discussion a section discussing the study limitations due to the study design (cross-sectional nature, lack of longitudinal data, presence of a distortion due to the type of administration/assistance, etc).

Response:
According to the suggestion, we have added a section to discuss the limitations of this study due to the study design.

Reviewer: Dr. Ute Ellert

1. Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Response:
Two native English-speaking colleagues helped to revise the paper.

2. page 3, section background, 2nd break: reference 2 doesn’t seem to be adequate

Response:
This reference was deleted.

3. The term ‘risk factor’ is not specified clearly. Is it only about factors which have a negative influence on quality of life? In particular in the chapter ‘Risk factors’ in the results section, it is not clear in which direction the described risk factors operate. In table 4 it is not clear, for example, which gender influences the quality of life. This should be verbalized more clearly.

Response:
In the revised manuscript, we provide more detailed explanation about the risk factors and how the variables were coded in Table 4.