Science Letters:

Crossover randomized controlled trial of the electronic version of the Chinese SF-36[*]

CHEN Tian-hui[1,2], LI Lu[†‡1], SIGLE Joerg M.[3], DU Ya-ping[1], WANG Hong-mei[1], LEI Jun[4]

([1]*Institute of Social Medicine and Family Medicine, Zhejiang University, Hangzhou 310058, China*)

([2]*Department of Medical Psychology and Medical Sociology, University of Goettingen, Goettingen 37073, Germany*)

([3]*Scientific IT Consulting, Duderstadt 37115, Germany*)

([4]*Institute for Informatics, University of Goettingen, Goettingen 37083, Germany*)

[†]E-mail: lilu@zju.edu.cn

Abstract: Objectives: to verify the feasibility and reliability of the electronic version of Chinese SF-36 based on the Quality-of-Life-Recorder. Design: A crossover randomized controlled trial, comparing a paper-based and an electronic version of the Chinese SF-36, was conducted. According to generated random numbers, interviewees were asked to fill out either the electronic version or the paper version first. The second version was filled in after a pause of at least 10 min. Settings and participants: One group of 100 medical students at the School of Medicine of Zhejiang University and the other group of 50 outpatients at a clinic for general practice in Hangzhou City (China) were eventually recruited in this study. Results: The acceptance of the electronic version was good (60% of medical students and 84% of outpatients preferred the electronic version). At the level of eight-scale scores, the mean-difference for each scale (except for general health) between the two versions was less than 5%. At the level of 36 questions, the percentage of "exact agreement" ranged within 64%~99%; the percentage of "global agreement" ranged within 72%~99%; 77% of the kappa coefficients demonstrated "good/excellent agreement" and 23% of the kappa coefficients demonstrated "medium agreement". Conclusion: This study, for the first time, can provide empirical basis for the confirmation of the feasibility and reliability of the electronic version of the Chinese SF-36 and may provide an impulse towards widespread deployment of the Quality-of-Life-Recorder in Chinese populations.

Key words: Health-related quality of life (HRQoL), SF-36, Electronic questionnaire, Computer based testing, General practice, Randomized control trial, Feasibility, Reliability

doi:10.1631/jzus.2007.B0604   **Document code:** A   **CLC number:** R395

INTRODUCTION

Over recent years, the SF-36 has become an important outcome measurement tool for health service research and clinical trials, especially for chronic diseases and cancer (Ware and Sherbourne, 1992). Traditionally, the SF-36 questionnaire is administered through paper versions and this data collection procedure infers costs with regard to time and resources.

In addition, the scoring method of SF-36 questionnaire is complicated. We feel that the tasks associated with data collection and the difficulties caused by the complicated scoring method might be major obstacles hindering widespread application of the paper version of the SF-36 questionnaire.

Along with the development of computer technologies, the technique of electronic data collection can dramatically reduce required time and costs, and a proper man-machine interface can still facilitate the completion of a survey for interviewees (Velikova *et al.*, 1999; Ryan *et al.*, 2002; Caro Sr *et al.*, 2001; Drummond *et al.*, 1995; Wilson *et al.*, 2002). In some European and American countries, the technique of

electronic data collection has already been applied in health-related quality of life (HRQoL) assessment. To the best of our knowledge, there is no related research on the technology of electronic data collection in the field of HRQoL research in China. Therefore, the conduction of research in China appears warranted both by the shortage of related research and the advantages of the technique of electronic data collection.

The Quality-of-Life-Recorder (QL-Recorder) is a platform for electronic recording of HRQoL data, which was developed by Dr. Sigle J.M., one of our coauthors (Sigle, 1994; 1995; Sigle and Porzsolt, 1996), and has similar advantages like the above mentioned technique for electronic data collection.

The research work presented in this paper is based upon a validated Chinese translation of the SF-36 developed by Dr. Lu Li, one of our coauthors (Li *et al.*, 2003). We used both the resulting paper version and an electronic version, generated by transferring the Chinese SF-36 onto the QL-Recorder.

In this article, we firstly intend to confirm the feasibility of the electronic version by evaluating the acceptance of the electronic version. Secondly, we aim to verify the equivalence of results derived with paper and electronic version by comparing the score equivalence between the two versions in a crossover randomized controlled trial. Since the psychometric performance (validity and reliability) of the paper version of the Chinese SF-36 had already been verified by Li *et al.*(2003), it may serve as a standard for the assessment of the electronic version. We believe that this study, for the first time, can provide an empirical basis for the confirmation of the feasibility and reliability of the electronic version of the Chinese SF-36 and may provide stimulus towards its widespread deployment.

METHODS

Design

Motivated by Ryan *et al.*(2002), a crossover randomized controlled trial was performed in this study. The detailed procedure is shown in Fig.1. First, the interviewees were asked to complete a demographic questionnaire that including questions on age, gender, education and frequency of computer usage. Second, a generated random number was used to assign interviewees to two groups and interviewees were asked to fill out either the electronic version or the paper version first. After a pause (at least 10 min) designed to reduce the effect of short term memory, interviewees were asked to complete the other version. After completing the two versions, the interviewees were asked which one they prefer. The duration required to complete the paper version was manually recorded; while the duration required to complete the electronic version was recorded automatically.
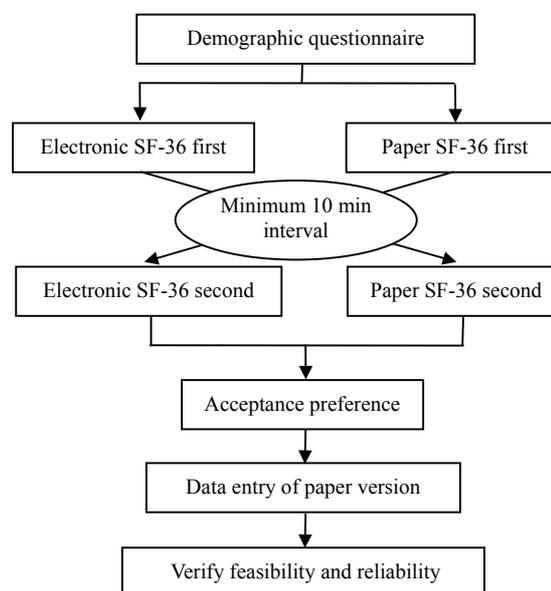

Fig.1 Research design

To detect a small systematic difference of 5% (a five-point difference in the transformed scale score that has a range of 100) in the SF-36 scores between the two methods of administration in a crossover randomized controlled trial with a power of 80% and 5% two-sided significance level, a required sample size of 150 interviewees was determined.

Setting and participants

This study used a convenience sample selected from a large range of age levels and different computer experiences. In our experiment, the convenience sample mainly consisted of two groups: One group included medical students with good computer experiences; while the other group included

outpatients from a center of community health services, including a large range of age levels (outpatients who could not read and with bad compliance should be excluded).

One group of 100 medical students at the School of Medicine of Zhejiang University and the other group of 50 outpatients at a clinic for general practice in Hangzhou City (China) were eventually recruited in this study.

Survey method

The first author acted as a surveyor and conducted all face-to-face surveys, in which interviewees filled out the questionnaires by themselves. Before the survey, the surveyor gave a brief introduction about the electronic version. During the survey, the surveyor would offer necessary help related to the electronic version whenever interviewees needed it.

Statistical methods

Statistical analysis was performed using QL-Recorder software and the Statistical Package for Social Science (SPSS 13.0 version for Windows).

RESULTS

Demographic characteristics and version preferences of the sample subjects

We found statistically significant difference between medical students and outpatients in "educational level" and "computer experience" (both of $P<0.001$, Table 1) and also found statistically significant difference between medical students and outpatients with regard to "preference of either version" ($P<0.05$). In the group of medical students, 94% reported "use computer frequently", but only 60% selected "prefer electronic version" and 25% selected "do not care"; whereas in the group of outpatients, only 38% reported "use computer frequently" and 24% reported "never use computers", however, 84% of outpatients selected "prefer electronic version" and 12% "do not care". Thus, we found that there was no positive correlation between "use computers frequently" and "like electronic version". Furthermore, the result indicated that the electronic version had good acceptances among both groups.

Table 1 Demographic characteristics and version preferences of the 150 subjects

	Medical students	Outpatients	χ^2	P
Sample quantity	100 (66.7%)	50 (33.1%)		
Sex			3.42	>0.05
Male	54 (54%)	19 (38%)		
Female	46 (46%)	31 (62%)		
Average age (year)	23.0	46.4	−	<0.001
Education			52.80[a]	<0.001
High school	2 (2%)	24 (48%)		
University	75 (75%)	25 (50%)		
Post-graduate	23 (23%)	1 (2%)		
Computer experience			61.04	<0.001
Often use	94 (94%)	19 (38%)		
Rare use	6 (6%)	19 (38%)		
Never use	0 (0%)	12 (24%)		
Preference to			9.108	<0.05
Paper version	15 (15%)	2 (4%)		
Electronic version	60 (60%)	42 (84%)		
Do not care	25 (25%)	6 (12%)		

[a]Fisher

Comparisons regarding the quality and the completeness of data

Compared with the electronic version, disadvantages of the paper version became obvious: A number of answers (110 questions) were missing, and other problems regarding data quality occurred, such as selection of multiple answers or ambiguous answers (affecting 2.5% of the answers to totally 5400 questions in the paper version). In addition, it took 5 h to check data quality and another 5 h to enter data into computers.

With regard to the electronic version, it could provide high-quality data and ensure 100% data completeness. No unintentionally missing data occurred in the electronic version since the QL-Recorder software did not allow unintentionally unanswered question. Furthermore, the electronic version allowed neither multiple nor ambiguous answers, e.g., selecting two choices.

Completion time for the two versions

Table 2 shows that, in an overall analysis, the electronic version was completed much more quickly than the paper version. The mean-difference of completion time was 208 s vs 255 s, which was of statistical significance ($P<0.001$, 95% CI: −64~−30 s).

An analysis taking into account the sequence, with which both versions were applied, showed that the 51 interviewees who completed the electronic version before filling in the paper version, required longer for the electronic version than for the paper version. The mean-difference here was 55 s with statistical significance ($P<0.001$, 95% *CI*: 36~73 s). The 99 interviewees who completed the paper version before filling in the electronic version, required longer time for the paper version. The mean-difference here was 99 s with statistical significance ($P<0.001$, 95% *CI*: −116~−83 s). This shows that the first completion of either version took longer than the second completion of the questionnaire, probably as interviewees' were not familiar with the content of the questionnaire when they filled it in the first time.

An analysis taking into account the different groups of participants showed that medical students completed the electronic version much quicker than the paper version. The mean-difference was 63 s with statistical significance ($P<0.001$, 95% *CI*: −83~−43 s). Outpatients also completed the electronic version quicker than paper version. The mean-difference was 15 s without statistical significance ($P>0.05$, 95% *CI*: −47~78 s).

Analysis of score equivalence

Two aspects of score equivalence were addressed. The first analysis was carried out at the level of eight-scale scores and the second analysis was carried out at the level of 36 questions.

Level of eight-scale scores

The mean-difference at the level of eight-scale scores was defined as the mean of electronic version minus that of paper version ($E−P$). Table 3 shows that the mean-difference for each scale, except for general health scale, between the two versions was less than 5.00 (5%), and not statistically significant ($P>0.05$); while the mean-difference for the scale of general health between the two versions was 2.33 (2.33%) with statistical significance ($P<0.05$, 95% *CI*: 0.91~3.74).

Level of 36 questions

According to Velikova *et al.*(1999), the agreement of the level of 36 questions between the two versions should also to be considered. The percentage of "exact agreement" ranged within 64%~99%; whereas the percentage of "global agreement" within 72%~99%. In addition, kappa coefficients for the 36 questions ranged within 0.53~0.89 (7 were between

Table 2 Completion time for the two versions of the SF-36, mean±*SD*

	N	Electronic version (s)	Paper version (s)	Mean-difference (s)	*t*-test	*P* value	95% *CI* (s)
Total time	150	208±89	255±77	−47	−5.42	<0.001	−64~−30
Version order							
Electronic first	51	275±99	221±66	55	5.96	<0.001	36~73
Paper first	99	173±59	273±77	−99	−12.01	<0.001	−116~−83
Group							
Medical student	100	178±69	241±72	−63	−6.38	<0.001	−83~−43
Outpatients	50	267±95	282±81	−15	−0.92	>0.05	−47~18

Table 3 SF-36 score's mean-difference at the level of scale scores[a]

Scale	Electronic version	Paper version	Mean-difference	*t*-test[b]	*P* value	95% *CI*
PF	92.23±13.57	92.47±12.86	−0.23±9.11	−0.31	>0.05	−1.70~1.24
RP	85.33±29.14	82.67±32.51	2.67±16.68	1.96	>0.05	−0.02~5.36
BP	78.11±18.16	77.10±19.40	1.01±9.96	1.24	>0.05	−0.60~2.61
GH	68.67±16.30	66.34±16.79	2.33±8.76	3.25	<0.05	0.91~3.74
VT	66.33±14.28	65.30±13.78	1.03±8.83	1.43	>0.05	−0.39~2.46
SF	82.92±14.87	84.25±14.80	−1.33±11.60	−1.41	>0.05	−3.20~0.54
RE	79.78±34.08	80.44±34.15	−0.67±27.43	−0.30	>0.05	−5.09~3.76
MH	71.01±12.92	69.92±13.06	1.09±8.09	1.66	>0.05	−0.21~2.40

[a]Score "100" means best HRQoL; [b]Paired *t*-test. PF: Physical function; RP: Role limitations due to physical problems; BP: Body pain; GH: General health; VT: Vitality; SF: Social functioning; RE: Role limitations due to emotional problems; MH: Mental health

0.41~0.60; 21 were between 0.61~0.80; 2 were between 0.81~0.100; and the other 6 could not be calculated according to the definition of the kappa coefficient because the above mentioned six questions had no answers within some categories), i.e., 77% of 30 available kappa coefficients ranged within 0.61~1.00, while 23% kappa coefficients within 0.41~0.60. According to the work of Landis and Koch (1977), 77% of the kappa coefficients thus fell into the category of "good/excellent agreement" and 23% of the kappa coefficients fell into the category of "medium agreement".

DISCUSSION

Three related aspects were addressed in this paper: First, the advantages of the electronic version were verified: e.g., saving time; high data quality; direct and quick availability of computed results; and convenient interface for data transfer into statistical analysis software. Second, we verified the feasibility of the electronic version: the acceptance of the electronic version was good (60% of medical students and 84% of outpatients preferred the electronic version, and another 25% and 12%, respectively, expressed no preference). Third, we verified the test/re-test reliability using a crossover application of the electronic and paper versions and comparing results at two levels of analysis.

In terms of data quality, no missing data occurred in the electronic version, reproducing the findings of Sigle and Porzsolt (1996) with a completeness rate of the electronic version surpassing 99.96% and missing data not coming from the overlooked questions or invalid answers, but from the (rare) unwillingness of individual outpatient to answer certain questions. In contrast, selection of multiple answer fields or ambiguous marks occurred in the paper version (affecting 2.5% of the answers to totally 5400 questions). This result also matches findings of previous studies (Velikova *et al.*, 1999; Ryan *et al.*, 2002), where missing data, multi-selection and ambiguous answers affected 5% to 10% of answers collected with a paper version.

It also confirms the advantages of electronic measurement with the QL-Recorder software demonstrated by Sigle (1995)'s and Holch (2000)'s doctoral dissertations.

CONCLUSION

We believe that this study, for the first time, can provide empirical basis for the confirmation of the feasibility and reliability of the electronic version of the Chinese SF-36 and may provide an impulse towards widespread deployment of the QL-Recorder in Chinese populations.

References

Caro Sr, J.J., Caro, I., Caro, J., Wouters, F., Juniper, E.F., 2001. Does electronic implementation of questionnaires used in asthma alter responses compared to paper implementation? *Qual. Life Res.*, **10**(8):683-691.

Drummond, H.E., Ghosh, S., Ferguson, A., Brackenridge, D., Tiplady, B., 1995. Electronic quality of life questionnaires: a comparison of pen-based electronic questionnaires with conventional paper in a gastrointestinal study. *Qual. Life Res.*, **4**(1):21-26. [doi:10.1007/BF00434379]

Holch, S., 2000. Practical Aspects of Standard Measurement of Quality-of-Life for In-Patient with an Electronic Quality-of-Life-Recorder. M.D. Thesis, University of ULM (in German).

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, **33**(1): 159-174. [doi:10.2307/2529310]

Li, L., Wang, H.M., Shen, Y., 2003. Chinese SF-36 health survey: translation, cultural adaptation, validation, and normalisation. *J. Epidemiol. Community Health*, **57**(4): 259-263. [doi:10.1136/jech.57.4.259]

Ryan, J.M., Corry, J.R., Attewell, R., Smithson, M.J., 2002. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Qual. Life Res.*, **11**(1):19-26. [doi:10.1023/A:1014415709997]

Sigle, J., 1994. Quality of Life: Computer Assistant Measurement of Quality of Life. QL-Recorder Software Manual. Kunstvolle EDV & Elektronik (in German).

Sigle, J., 1995. Practical Aspects of Quality-of-Life Measure: Standard Measurement of Quality-of-Life for Out-Patients with an Electronic Quality-of-Life-Recorder. M.D. Thesis, University of ULM (in German).

Sigle, J., Porzsolt, F., 1996. Practical aspects of quality-of-life measurement: design and feasibility study of the quality-of-life recorder and the standardized measurement of quality of life in an out-patient clinic. *Cancer Treat. Rev.*, **22**(Suppl. A):75-89. [doi:10.1016/S0305-7372(96)90067-5]

Velikova, G., Wright, E.P., Smith, A.B., Cull, A., Gould, A., Forman, D., Perren, T., Stead, M., Brown, J., Selby, P.J., 1999. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. *J. Clin. Oncol.*, **17**(3):998-1007.

Ware, J.E., Sherbourne, C.D., 1992. The MOS 36-Item Short Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med. Care*, **30**(6):473-483. [doi:10. 1097/00005650-199206000-00002]

Wilson, A.S., Kitas, G.D., Carruthers, D.M., Reay, C., Skan, J., Harris, S., Treharne, G.J., Young, S.P., Bacon, P.A., 2002. Computerized information-gathering in specialist rheumatology clinics: an initial evaluation of an electronic version of the Short Form 36. *Rheumatology*, **41**(3): 268-273. [doi:10.1093/rheumatology/41.3.268]