國立臺灣大學公共衛生學院

職業醫學與工業衛生研究所

碩士論文

比較個案自填方式不同對於生活品質問卷再測信度之影響

A comparison of test-retest reliability of self-reported SF-36, WHOQOL, and EQ-5D questionnaires based on different administration approaches

研究生：林育任 撰

指導教授：王榮德 陳健弘 博士

中華民國 九十八 年 五 月

May, 2009

目錄

摘要 3

ABSTRACT 5

INTRODUCTION 7

MATERIALS AND METHODS 9

PATIENTS SELECTION 9

HRQOL QUESTIONNAIRES 9

WHOQOL-BREF 9

SF-36 10

EQ-5D 10

DATA COLLECTION 11

STATISTICAL ANALYSIS 11

RESULTS 12

DISCUSSION 15

REFERENCE 18

圖表目錄

Table 1 Clinical and demographical data of chronic liver disease..... 22

Table 2 Intra-class correlation coefficient and 95%CI (SF-36, WHOQOL-BREF, and EQ-5D in two approaches) 23

Table 3 Mean difference of SF-36, WHOQOL-BREF, EQ-5D in two approaches……...24

Table 4 Multiple linear regression model for determinants associated with two approaches (N=179) among SF-36, WHOQOL-BREF and EQ-5D……..25

Table 5 Table 5 Comparison of the floor and ceiling of the SF-36, WHOQOL-BREF, and EQ-5D of pre-test in hospital interview and mail surveys……29

Table 6 Table 6 Comparison of the floor and ceiling of the SF-36, WHOQOL-BREF, and EQ-5D of pre-test in two mail surveys…….30

摘要

目的：

本研究目的為使用 SF-36, WHOQOL-BREF, 和 EQ-5D 三種國際間認可且常用之生活品質問卷，探討慢性肝病病人是否會因為個案收案的方式不同而導致再測可信度測試的結果有所差異。

方法：

本研究於台灣北部某大型教學醫院，於門診時收集慢性肝病病人自填式的生活品質問卷，其中採取兩種不同的收案方式進行，第一種方式：先於醫院門診填寫前測問卷，於兩週後讓個案在家中填寫後測問卷，再以郵寄方式寄回。第二種收案方式：則是前測與後測問卷都是讓個案帶回家填寫後以郵寄方式，寄回間隔也是兩週。本研究經過台大醫院研究倫理委員會審查通過後進行。第一種方式同意參與的個案人數為 69 人，完成前測與後測兩次問卷填寫之有效問卷共 52 份 (回覆率=75.4%)。第二種方式同意參與的個案為 180 人，完成前測與後測兩次問卷填寫之有效問卷共 127 份 (回覆率=70.6%)。問卷分數經計算後，以 paired-*t* test 和 intra-class correlation coefficient (ICC)比較三種問卷再測信度，及使用 Bootstrap re-sampling 的方式檢定兩種收案方式的再測信度是否具統計上的差異。使用 independent *t* test 比較兩種收案方式之前後分數差異，再以多迴歸分析控制相關潛在性干擾因子進行分析探討。

結果：

第一種收案方式問卷再測信度 ICC 值最低為 0.37 最高達 0.81；第二種收案方式則為 0.61-0.87，經分析比較後發現這三份問卷在第二種收案方式中各面向的 ICC 分數均高於第一種收案方式。在 Independent *t* test 兩種不同收案方式的比較上，除了 EQ-5D 疼痛狀態有達統計上顯著差異以外 (分別為 0.1 ± 0.4 vs. 0.0 ± 0.3, p=0.04)，其他問卷中狀態變項皆無發現有所差異，並進一步使用多迴歸分析，在控制年齡、性別、教育程度後，並無統計上的顯著差異（p=0.06）。比較前測的三種問卷後，發現 EQ-5D 有較大的天花板效應(75.0%-100.0%)而且平均 utility 值很高(0.95±0.08)，證明這個問卷有較少的敏感性去量測疾病程度較輕的狀況，

WHOQOL-BREF 產生的天花板效應較小（1.9%）而且沒有地板效應。

結論：

本研究的結果發現以郵寄方式收案及使用 WHOQOL-BREF 和 SF-36 問卷較適合慢性肝病病人的生活品質調查，建議未來針對此族群進行生活品質研究者，可將本研究結果當作參考。

關鍵字：訪視者效應 生活品質 填寫方式

Abstract

OBJECTIVES: This aim of this study is to examine whether, and to what extent the test-retest reliability of Short-Form-36 (SF-36), World Health Organization Quality of Life (WHOQOL-BREF), and EuroQol-5D (EQ-5D) questionnaires are affected by the administration approaches in patients with chronic liver disease.

METHODS: Patients with chronic liver disease were recruited from the outpatient department of a medical center in Taiwan. Their self-reported questionnaires were collected by two approaches. The first approach is that patients received an interview and filled the questionnaire in hospital. Retest questionnaires were returned by mail two weeks later. The other approach is that patients filled both test and retest questionnaires at home, under researcher's instruction, during the two-week period. Of 69 patients recruited for the first approach; 52 persons completed both questionnaires (75.4%), while the response rate of the second approach is 70.6% (127 of 180). The response rate in the second approach was better (two mail surveys) than that in the first approach (interview at the out patient clinics followed by mail).

After scoring questionnaires, a paired-*t* test and intra-class correlation coefficient (ICC) were conducted to compare test-retest reliability for three questionnaires. The difference of mean score between two approaches was examined by independent *t* test. Analyses of mean score differences of different domains were performed by multiple linear regressions with different administration methods, age, gender, and education as covariates.

RESULTS: The test-retest reliabilities ranged from 0.37 to 0.81 in the first approach, while they ranged 0.61-0.87 in the 2nd approach. Most ICCs were higher in the 2nd approach than the 1st approach. There was also no significant difference in the mean difference of test-retest results in two approaches, except in the dimension of 'pain/discomfort' mean difference (0.1±0.4 and 0.0 ± 0.3, p=0.04) by EQ-5D but the result were not found by multiple linear regression, after controlling variables of age, sex, and education (p=0.06). Among the three questionnaires, the large ceiling effect (75.0%-100.0%) and the high mean score for utility (0.95±0.08) of the EQ-5D suggest it could be less responsive to detect any difference when the severity of illness is low. In contrast, the scores of WHOQOL-BREF have wider ranges, less ceiling effect （1.9%） and no floor effect.

CONCLUSIONS: The high reliabilities of the present study supported that mail survey method is acceptable for measuring quality of life, and the WHOQOL-BREF and the SF-36 questionnaires of instrument could be used for assessing HRQoL in patients with asymptotic carrier.

Key words: interviewer's effect, Quality of life, mode of administration

INTRODUCTION

The infection of hepatitis B virus (HBV) and hepatitis C virus (HCV) is a serious issue in Taiwan. The seroprevalence of HBV and HCV has been 15% to 20% and 2% to 4% of the Taiwanese population respectively (1, 2). For this group of people, their risk of developing hepatocellular carcinoma (HCC) is higher than the normal population. Chronic HBV and HCV infections are major etiologic factors of liver cirrhosis and HCC in Taiwan (3), and lead to about 12,000 liver-related deaths every year.

Since past decade, health-related quality of life (HRQoL) had been increasingly employed in assessing the cost-effectiveness of medical cares (4-8). By using this indicator, studies that evaluated the quality of life among patients with chronic hepatic disease have reported that these patients had lower quality of life (9-11). The potential determinants of quality of life included severity of disease, depression, fatigue, and loss of appetite and etc (12). There have been many questionnaires designed to assess quality of life, such as Short-Form-36 (SF-36), World Health Organization Quality of Life (WHOQOL-BREF), or EuroQol-5D (EQ-5D) (13, 14). For chronic liver diseases, most studies used SF-36, Short-Form-6D (SF-6D) or Short-Form-12 (SF-12), Health Utilities Index (HUI) as the measurement tools (15, 16). In the study of Lee et al. (17) WHOQOL-BREF questionnaire has been demonstrated as an effective measurement tool to investigate the quality of life of patients with hepatic cancer. Although EQ-5D was seldom used to evaluate the HRQoL in patients with chronic liver disease, it provides a simple descriptive profile and a single index value for health status that can be used in cost-effectiveness evaluation.

There are various approaches to conduct quality of life study, including mail, telephone, and face-to-face interview. In general, face-to-face interview is relatively

time- and labor-consuming but with minimum missing data (18). Although mail investigation might be much cheaper, the response rate is usually low (19, 20). Till now, it is still controversial which method is the best one to study quality of life for patients with hepatic disease. Therefore, this study aims to examine whether, and the extent to which, the test-retest reliability of SF-36, WHOQOL-BREF, and EQ-5D questionnaires are affected by the administration approaches in patients with chronic liver disease.

MATERIALS AND METHODS

Patients Selection

Patients with chronic liver disease were recruited from the outpatient of a medical center in Taiwan from August 2006 to October 2006. Patients who have other malignancy, consciousness unclear and illiterate problems were excluded. All subjects were inactive HBV carrier, i.e. their serum levels of ALT were normal and there were no evidences of liver cirrhosis by abdominal ultrasound and gave written informed consent. The study was granted by the Institutional Review Boards of the National Taiwan University Hospital before commencement.

HRQoL Questionnaires

WHOQOL-BREF

The World Health Organization Quality of Life-BREF Taiwan version consists of four domains - physical health, psychological, social relationships, and environment. The WHOQOL-BREF Taiwan version contained 28 items, including 26 original items same as in the WHOQOL-BREF and another two Taiwanese-specific items. In particular, these two Taiwanese-specific items are "Do you feel respected by others?" and "Can you always eat as you wish?" in social (TW), and environment (TW) domains. Previous studies had indicated that the reliability and validity of Taiwanese WHOQOL-BREF were well developed (21). The domain scores range from 0 to 100, with a higher score indicating a better QOL on the corresponding domain.

SF-36

The SF-36 Taiwan Version questionnaire validated by Lu and et al. (13, 22) was self-administered. In SF-36, 36 items were developed to measure 8 quality- of-life concepts: physical functioning (PF); social functioning (SF); role limitations-physical (RP); role limitations-emotional (RE); mental health (MH); vitality (VT); bodily pain (BP); and general health perception (GH), which could be aggregated into two summary scores: a physical component summary (PCS) and a mental component summary (MCS). Scores are sub-scaled and summarized, ranging from 0 to 100. PCS and MCS were calculated using the data from the general Taiwan population (23). The high scores represented that the people had good health status.. When there were items being missed in patients' self-report, they would be replaced by the mean score from the respondent's remaining items (24).

EQ-5D

EQ-5D was a standardized measure of health status developed by the EuroQoL Group in order to provide a simple, generic measure of health for clinical and economic appraisal (25). The EQ-5D Taiwan Version questionnaire was self-administered. Five dimensions included - mobility, self-care, usual activities, pain/discomfort, and anxiety/ depression. Patients might choose one level, designated as 1 to 3, that reflects their health status in each dimension. The high scores represented that the people had bad health status. (26). The health states could be weight to a single summary index that was obtained from a random sample of the adult population in the United Kingdom in EQ5D. The index spanned a scale from 0 (as bad as being dead) to 1.0 (perfect health) (27).

Data Collection

We used randomly consecutive allocation to enroll our subjects. Two approaches were employed to collect patients' self-reported questionnaires. The first approach was that patients received a self-completion in the presence of an interview and filled the questionnaire in hospital. Retest questionnaires were returned by mail two weeks later. The other approach is that patients filled both test and retest questionnaires at home, under researcher's instruction, during two-week period.

Statistical Analysis

After scoring questionnaires, a paired-*t* test and intra-class correlation coefficient (ICC) were used to compare test-retest reliability for three questionnaires. The ICC is an index of concordance for dimensional measurements ranging between 0 and 1, where $\geqq 0.7$ is considered good reliability and $\geqq 0.5$ is acceptable (28, 29). We need to compare the agreement of ICC with bootstrap re-sampling between two approaches in survey. From the original data set, the bootstrap generates a large number of samples and estimates the sampling distribution of a statistic such as mean and variance (30).We compared thirty times and fifty times on bootstrap sampling. There was no statistically significant difference between two different times. So we used thirty times for bootstrap sampling. Mean difference between two approaches was examined by independent *t* test. Analyses of mean differences of different domains were performed by multiple linear regressions taking age, gender, level of education, and different modes of administration as independent variables. Differences in categorical values were tested by χ^2 or Fisher's exact test, with a *p* value of less than 0.05 considered as statistically significant. All statistical analyses were performed using SAS 9 (SAS Institute, Cary, NC).

RESULTS

Subject characteristics

There are 220 patients with chronic liver disease recruited based on the criteria in this study. Of 69 patients recruited for the first approach, 52 (75.4%) persons completed both questionnaires. In contrast, the response rate of the second approach was 70.6% (127 of 180). On average, the patients in first approach were 45.9 years old and 73.1% were men. 42 patients were infected with HBV; eight patients had positive anti-HCV; one patient with both HBV and HCV; and one patient without HBV or HCV infection. Patients recruited in the second approach on average were 48.5 years old, in that 59.8% were men. A total of 86.6% patients were with HBV. Summary statistics are shown in Table 1. There is no statistically significant difference in age, gender, education, etiology of disease and presence or absence of cirrhosis among two groups of patients.

Test-retest reliability

The intra-class correlation coefficient (ICC) was calculated to estimate the test-retest reliability of each scale of the SF-36, each domain of WHOQOL-BREF, and the utility score of EQ-5D. As shown in Table 2, the ICC of SF-36 ranged from 0.37 to 0.81 in the first approach survey and from 0.61 to 0.87 in the second approach. The lowest ICCs were observed for social function, role limitations physical, and role limitations emotional. With respect to WHOQOL-BREF, the ICC ranged from 0.71 to 0.84 in four domains. The ICC of EQ-5D was 0.78 and 0.61 for 1st and 2nd approach respectively.

Except for three items, most ICCs were higher in the 2nd approach. There were no statistically significant differences in the ICC in role limitations physical of SF-36 and

in physical domain of WHOQOL-BREF between these two approaches. The ICCs were higher in the 1st approach in bodily pain of SF-36 and in the utility of EQ-5D.

In the test-retest mean difference results, there were also no significant difference among three questionnaires in two approaches, except in the dimension of 'pain/discomfort' mean difference (0.1±0.4 and 0.0 ± 0.3, p=0.04) by EQ-5D but the result were not found by multiple linear regression, after controlling age, sex, and education (p=0.06) (Table 3 and Table 4).

Descriptive statistic for the SF-36, the WHOQOL-BREF, and the EQ-5D

In the 1st test of hospital interview and mail surveys, the scores ranged from 56.1±20.6 to 91.5±9.2 on eight domains of the SF-36, 60.4±13.8 to 69.7±12.7 for the WHOQOL-BREF, and 0.95±0.06 for the utility of EQ-5D. In the 1st test of two mail surveys, the scores ranged from 60.9±20.7 to 91.8±14.1 on eight domains of the SF-36, 62.4±13.0 to 73.6±11.1 for the WHOQOL-BREF, and 0.95±0.08 for the utility of EQ-5D (Table3).

Based on the scores of our studies, it sounds that a large ceiling effect was observed for the EQ-5D dimensions. The percentage at the 'ceiling' of the functional dimensions was over 94% for the EQ-5D, while it was to35-58% for the SF-36 and 2% for the WHOQOL-BREF, and on emotional well-being 75% for EQ-5D (anxiety/ depression) and 4% for SF-36 (mental health) and on pain 81% for EQ-5D compared to 37% on SF-36. The flooring effects were observed for two the role-disability scales (15% for RP and 21% for RE) in the SF-36 but there were no floor in four domains of the WHOQOL-BREF and the EQ-5D dimensions. The similar results were observed in

the two mail surveys (Table 6).

DISCUSSION

Liver disease imposed a great burden on Taiwanese society. Treatments in patients with chronic liver disease aimed to reduce complications and improved survival. Clinician used biochemical tests, virological tests and image studies to evaluate the severity of chronic liver disease. However, the HRQoL is reduced in patients with chronic liver disease (9-11). Therefore, in addition to biochemical, virological and image parameters, measuring HRQoL was becoming important.

To successfully implement the HRQoL study on patients with chronic liver diseases, it is important to choose the most feasible mode of administration. The asymptomatic HBV carriers are followed up at an interval of 6-12 months. It was difficult to repeat measurement at short interval by face-to-face interview. Besides, the major reasons of patients' refusal to answer the questionnaires were lack of time and/ or the inconvenience imposed by an extra trip to the hospital. Thus, administration approaches other than face-to-face interview should be explored. We compared the test retest reliability between two strategies (hospital interview and mail surveys and two mail surveys). Our results showed that there were statistically significant differences between two approaches. The result suggested that mail survey may be an appropriate method.

The response rate was better in the first approach (interview at the out patient clinics followed by mail) than that in the second approach (two mail surveys) (75.4% v.s 70.6%). In previous studies, the response rate was lower in mail survey than telephone survey or face-to face survey (19, 20). The past studies showed that a mixed-mode survey strategy had a higher response rate than either a mail or telephone survey alone

(31-34). Our result was consistent with these studies.

In addition to the response rate, the results of ICC also supported that the 2nd approach was better than the 1st approach. In the 2nd approach (two mail surveys), the ICC of all items ranged from 0.61-0.87, i.e., from accepted to good for reliability for three HRQoL questionnaires. Besides, the ICCs were higher in most of the items of the 2nd approaches than the 1st approach in SF-36 and WHOQOL-BREF questionnaires. This implied that the 2nd approach (two mail surveys) yielded better test-retest reliability than the 1st approach (interview + mail surveys).

However, the ICC score was lower in the utility of EQ-5D in 2nd approach. The possible reason for the result was that the utility of EQ-5D was reported on a particular day and those subjects may be have other diseases other than liver diseases that might influence the scores of EQ-5D. Additional research will be needed to prove our assumption.

The ICCs were low in social functioning, and role limitations emotional of the SF-36 questionnaire in both approaches. The finding was consistent with previous studies that intra-individual variation might explain why the score was low (18, 35). Similar findings were not found in social domain of WHOQOL-BREF questionnaire. The SF-36 asked subjects to report their health over the previous 4 weeks, while the WHOQOL-BREF focused on health status over the past 2 weeks. It was probably the longer recall periods in SF-36 that permits more opportunities for events to occur. Another explanation for the low score in social functioning was that there were only two items in that dimension.

Among the three questionnaires, the skewness of the distributions reflects the

limitation of having only one or two categories on EQ-5D, compared with a range of possible raw scores of 4-24 for the SF-36, 10-22 for the WHOQOL-BREF in hospital interview and mail surveys. The large ceiling effect and the high mean score for utility of the EQ-5D suggest EQ-5D might be less responsive to detect differences for minimal to mild diseases, such as inactive HBV carriers in our study. The scores of WHOQOL-BREF have wide range, less ceiling effect and no floor effect. Our result of flooring and ceiling effect for two role-disability scales in the SF-36 were consistent with what have been found in previous studies. These two scales are the coarsest of the SF-36 scales and are more susceptible to floor and ceiling effects than any of the other six scales (6). Taken together, the WHOQOL-BREF and the SF-36 are more sensitive instruments, particularly for patients with asymptotic carrier.

There were several limitations in this study. First, background information of non-respondents was not collected in this study. Therefore, we were unable to compare the characteristics between respondents and non-respondents. Second, we could not make sure whether questionnaires of test and retest were finished by the same respondent.

In conclusion, mail survey method, the WHOQOL-BREF and the SF-36 questionnaires of instrument are useful for assessing HRQoL in patients with asymptotic carrier.

.

REFERENCES

1. Chen, D. S., Kuo, G. C., Sung, J. L., Lai, M. Y., Sheu, J. C., Chen, P. J., Yang, P. M., Hsu, H. M., Chang, M. H., & Chen, C.J., et al. Hepatitis C virus infection in an area hyperendemic for hepatitis B and chronic liver disease: the Taiwan experience. J Infect Dis 1990;162:817-22.
2. Chen, C. H., Yang, P. M., Huang, G. T., Lee, H. S., Sung, J. L., & Sheu, J. C. Estimation of seroprevalence of hepatitis B virus and hepatitis C virus in Taiwan from a large-scale survey of free hepatitis screening participants. J Formos Med Assoc 2007;106:148-55.
3. Chen, C. J., Yu, M. W., & Liaw, Y. F. Epidemiological characteristics and risk factors of hepatocellular carcinoma. J Gastroenterol Hepatol 1997;12:S294-308.
4. Liou, H. H. Health related quality of life in adult patients with epilepsy compared with a general reference population in Taiwan. Epilepsy Res 2005;64:151.
5. Hsu, C., Wang, J. D., Hwang, J. S., Tien, H. F., Chang, S. M., Cheng, A. L., Chen, Y. C. & Tang, J. L. Survival-weighted health profile for long-term survivors of acute myelogenous leukemia. Qual Life Res 2003;12:503-17.
6. Hsiung, P. C., Fang, C. T., Chang, Y. Y., Chen, M. Y., & Wang, J. D. Comparison of WHOQOL-BREF and SF-36 in patients with HIV infection. Qual Life Res 2005;14:141-50.
7. Yang, S. C., Kuo, P. W., Wang, J. D., Lin, M. I., & Su, S. Quality of life and its determinants of hemodialysis patients in Taiwan measured with WHOQOL-BREF (TW). Am J Kidney Dis 2005;46:635-41.
8. Chang, C. Y., Hung, C. K., Chang, Y. Y., Tai, C. M., Lin, J. T., & Wang, J. D. Health-related Quality of Life in Adult Patients with Morbid Obesity Coming for

Bariatric Surgery. Obes Surg 2008.

9. Bayliss, M. S. Methods in outcomes research in hepatology: definitions and domains of quality of life. Hepatology 1999;29:3S-6S.

10. Foster, G. R., Goldin, R. D., & Thomas, H. C. Chronic hepatitis C virus infection causes a significant reduction in quality of life in the absence of cirrhosis. Hepatology 1998;27:209-12.

11. Bondini, S., Kallman, J., Dan, A., Younoszai, Z., Ramsey, L., Nader, F., & Younossi, Z. M. Health-related quality of life in patients with chronic hepatitis B. Liver Int 2007;27:1119-25.

12. Gutteling, J. J., de Man, R. A., van der Plas, S. M., Schalm, S. W., Busschbach, J. J., & Darlington, A. S. Determinants of quality of life in chronic liver patients. Aliment Pharmacol Ther 2006;23:1629-35.

13. Lu, J. F. R, Tseng, H. M., & Tsai, Y. J. Assessment of Health-related Quality of Life in Taiwan (I): Development and Psychometric Testing of SF-36 Taiwan Version. Taiwan Journal Public Health 2003;22:501-11.

14. Krabbe, P. F. M., Peerenboom, L., Langenhoff, B. S., & Ruers, T. J. M. Responsiveness of the generic EQ-5D summary measure compared to the disease-specific EORTC QLQ C-30. Qual Life Res 2004;13:1247-53.

15. Gutteling, J. J., de Man, R. A., Busschbach, J. J., & Darlington, A. S. Overview of research on health-related quality of life in patients with chronic liver disease. Neth J Med 2007;65:227-34.

16. Younossi, Z. M., Boparai, N., McCormick, M., Price, L. L., & Guyatt, G. Assessment of utilities and health-related quality of life in patients with chronic liver disease. Am J Gastroenterol 2001;96:579-83.

17. Lee, L. J., Chen, C. H., Yao, G., Chung, C. W., Sheu, J. C., Lee, P. H., Tsai, Y. J.,

& Wang, J. D. Quality of life in patients with hepatocellular carcinoma received surgical resection. J Surg Oncol 2007;95:34-9.

18. Weinberger, M., Oddone, E. Z., Samsa, G. P., & Landsman, P. B. Are health-related quality-of-life measures affected by the mode of administration? J Clin Epidemiol 1996;49:135-40.

19. Duncan, P., Reker, D., Kwon, S., Lai, S. M., Studenski, S., Perera, S., Alfrey, C., & Marquez, J. Measuring Stroke Impact With the Stroke Impact Scale: Telephone Versus Mail Administration in Veterans With Stroke. Med Care 2005;43:507.

20. Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. Am J Public Health 1979;69:238-45.

21. Yao, G., Chung, C. W., Yu, C. F., & Wang, J. D. Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. J Formos Med Assoc 2002;101:342-51.

22. Tseng, H. M., Lu, J. F. R.,& Tsai, Y. J. Assessment of Health-related Quality of Life in Taiwan (Ⅱ): Norming and Validation of SF-36 Taiwan Version. Taiwan Journal Public Health 2003;22:512-8.

23. Lin, C. H. Validation and Norm Establishing of the SF-36 Taiwan Version. Unpublished Master Thesis 2003.

24. Fuh, J. L., Wang, S. J., Lu, S. R., Juang, K. D., & Lee, S. J. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community. Qual Life Res 2000;9:675-83.

25. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990;16:199-208.

26. Chang, T. J., Tarn, Y. H., Hsieh, C. L., Liou, W. S., Shaw, J. W., & Chiou, X. G.

Taiwanese version of the EQ-5D: validation in a representative sample of the Taiwanese population. J Formos Med Assoc 2007;106:1023-31.

27. Dolan, P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095-108.

28. Deyo, R. A., Diehr, P., & Patrick, D. L. Reproducibility and responsiveness of health status measures. Control Clin Trials 1991;12:142s-58s.

29. Peter, M., & Fayers, D. M. Quality of Life, 2002.

30. Chernick, M. R. Bootstrap Methods: A Guide for Practitioners and Researchers, 2008.

31. McHorney, C. A., Kosinski, M., & Ware, J. E., Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. Med Care 1994;32:551-67.

32. Brogger, J., Bakke, P., Eide, G. E., & Gulsvik, A. Comparison of telephone and postal survey modes on respiratory symptoms and risk factors. Am J Epidemiol 2002;155:572-6.

33. O'Toole, B. I., Battistutta, D., Long, A., & Crouch, K. A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. Am J Epidemiol 1986;124:317-28.

34. Hochstim, J. R. A Critical Comparison of Three Strategies of Collecting Data from Households. Journal of the American Statistical Association 1967;62:976-89.

35. Weinberger, M., Nagle, B., Hanlon, J. T., Samsa, G. P., Schmader, K., Landsman, P. B., Uttech, K. M., Cowper, P. A., Cohen, H. J., & Feussner, J. R. Assessing health-related quality of life in elderly outpatients: telephone versus face-to-face administration. J Am Geriatr Soc 1994;42:1295-9.

Table 1 Clinical and demographical data in patients with chronic liver disease

	Hospital interview and mail surveys (N=52)	**Two mail surveys (N=127)**	***P*-value**
Age (years), Mean ± SD	45.9 ± 12.9	48.5 ± 13.0	0.224[a]
Gender, N (%)			0.094 [b]
Male	38 (73.1%)	76 (59.8%)	
Female	14 (26.9%)	51 (40.2%)	
Education, N (%)			0.460 [b]
≥Bachelor degree	28 (53.9%)	76 (59.8%)	
<Bachelor degree	24 (46.1%)	51 (40.2%)	
Etiology of disease, N (%)			0.571 [b]
HBsAg (+)	42 (80.8%)	110 (86.6%)	
Anti-HCV (+)	8 (15.4%)	14 (11.0%)	
HBV+HCV	1 (1.9%)	2 (1.6%)	
Non-B, Non-C	1 (1.9%)	1 (0.8%)	
Cirrhosis, N (%)			0.352 [b]
Yes	2 (3.9%)	11 (8.7%)	
No	50 (96.1%)	116 (91.3%)	

[a.] Independent-t test; [b.] Chi-square test

Table 2 Intra-class correlation coefficient and 95%CI (SF-36, WHOQOL-BREF, and EQ-5D in two approaches)

	Hospital interview and mail surveys (N=52)	Two mail surveys (N=127)	*P*-value[a]
SF-36			
Physical Functioning	0.66 (0.47-0.79)	0.77 (0.69-0.83)	0.002
Social Functioning	0.37 (0.11-0.58)	0.61 (0.49-0.71)	<0.001
Role Limitations Physical	0.67 (0.49-0.80)	0.61 (0.49-0.71)	0.062
Role Limitations Emotional	0.46 (0.22-0.55)	0.65 (0.54-0.74)	<0.001
Mental Health	0.72 (0.56-0.83)	0.83 (0.77-0.88)	<0.001
Vitality	0.81 (0.68-0.88)	0.82 (0.75-0.87)	0.044
Bodily Pain	0.76 (0.61-0.86)	0.70 (0.59-0.78)	<0.001
General health perception	0.70 (0.53-0.82)	0.87 (0.82-0.91)	<0.001
Physical Component Summary	0.73 (0.56-0.83)	0.76 (0.67-0.83)	0.043
Mental Component Summary	0.77 (0.63-0.86)	0.85 (0.79-0.89)	<0.001
WHOQOL-BREF			
Physical	0.71 (0.54-0.82)	0.73 (0.64-0.80)	0.353
Psychological	0.76 (0.61-0.85)	0.84 (0.79-0.89)	<0.001
Social (TW)	0.72 (0.56-0.83)	0.79 (0.71-0.85)	0.006
Environment (TW)	0.79 (0.66-0.87)	0.80 (0.73-0.85)	0.031
EQ-5D			
Utility	0.78 (0.64-0.86)	0.61 (0.49-0.71)	<0.001

[a,] Bootstrap re-sampling thirty times

Table 3 Mean difference of SF-36, WHOQOL-BREF, EQ-5D in two approaches

	Hospital interview and mail surveys (N=52)			Two mail surveys (N=127)			
	Test [a.]	Re-test [a.]	Mean difference [a]	Test [a.]	Re-test [a.]	Mean difference [a]	*P*-value
SF-36							
Physical Functioning	91.5±9.2	89.9±13.5	-1.4±9.5	91.8±14.1	90.1±17.8	-1.7±10.7	0.897
Social Functioning	76.7±19.5	77.8±19.1	0±21.7	82.4±14.6	83.5±15.7	1.4±13.4	0.673
Role Limitations Physical	71.2±38.8	73.0±39.3	1.5±31.8	83.1±32.2	82.2±31.6	-0.8±28.1	0.640
Role Limitations Emotional	67.9±40.7	70.6±38.7	2.0±41.3	79.7±34.9	80.2±35.3	0.3±29.3	0.791
Mental Health	64.9±14.3	65.8±19.1	0.8±12.6	66.8±17.5	66.4±18.3	0.0±10.5	0.668
Vitality	59.5±19.5	61.5±19.4	1.4±12.1	62.2±17.9	61.0±17.2	-0.7±10.6	0.257
Bodily Pain	77.6±20.9	80.0±21.7	2.4±14.6	83.1±20.4	84.8±16.9	2.0±14.4	0.897
General health perception	56.1±20.6	56.0±22.2	-1.0±16.5	60.9±20.7	61.4±21.4	1.1±10.7	0.400
PCS	50.4±9.2	50.3±10.6	-0.2±7.3	53.4±10.1	53.4±9.3	-0.2±6.6	0.961
MCS	47.3±10.3	48.2±12.2	0.6±7.6	49.1±10.7	49.5±11.3	0.5±5.9	0.880
WHOQOL-BREF							
Physical	69.7±12.7	68.8±14.0	-0.5±10.1	73.6±11.1	73.6±11.5	0.0±8.3	0.747
Psychological	60.4±13.8	59.3±16.0	-0.6±10.4	63.7±14.2	63.3±15.0	-0.3±8.2	0.848
Social (TW) [b]	62.7±11.6	60.6±12.8	-2.0±8.9	62.4±13.0	63.2±14.1	0.8±8.8	0.056
Environment (TW) [b]	64.0±13.5	63.1±14.4	-0.7±9.0	67.3±10.6	66.7±11.4	-0.6±6.9	0.937
EQ-5D							
Mobility	1.0±0.1	1.0±0.0	0.0±0.1	1.0±0.2	1.0±0.1	0.0±0.1	0.863
Self-care	1.0±0.0	1.0±0.0	0.0±0.0	1.0±0.1	1.0±0.0	0.0±0.1	0.319
Activities	1.1±0.2	1.1±0.3	0.0±0.3	1.0±0.2	1.0±0.1	0.0±0.2	0.554
Pain/ discomfort	1.2±0.4	1.3±0.5	0.1±0.4	1.2±0.4	1.2±0.4	0.0±0.3	0.036*
Anxiety/ depression	1.3±0.4	1.3±0.5	0.1±0.4	1.2±0.5	1.3±0.5	0.0±0.4	0.701
Utility	0.95±0.06	0.93±0.08	-0.02±0.05	0.95±0.08	0.95±0.07	0.00±0.07	0.095

[a.] Mean ± SD [b] Domain including culture-specific items. * $p<0.05$

Table 4 Multiple linear regression model for determinants associated with two approaches (N=179) among SF-36, WHOQOL-BREF and EQ-5D

Variable	*β*	SE	*p* Value
SF-36			
Physical Functioning Difference	R^2-adj = 0		
Intercept	0.6	4.0	0.877
Age	-0.1	0.1	0.438
Education (<Bachelor degree=0)	0.6	1.7	0.749
Gender (Male =0)	-0.2	1.7	0.905
Two approaches (Two mail surveys =0)	0.1	1.8	0.973
Social Functioning Difference	R^2-adj = 0		
Intercept	2.0	6.1	0.742
Age	-0.1	0.1	0.577
Education (<Bachelor degree=0)	2.3	2.7	0.406
Gender (Male =0)	2.0	2.6	0.431
Two approaches (Two mail surveys =0)	-1.2	2.8	0.671
Role Limitations Physical Difference	R^2-adj = 0		
Intercept	-3.5	11.0	0.751
Age	0.0	0.2	0.847
Education (<Bachelor degree=0)	2.0	4.9	0.691
Gender (Male =0)	-0.6	4.7	0.905
Two approaches (Two mail surveys =0)	2.4	5.0	0.626
Role Limitations Emotional Difference	R^2-adj =0		
Intercept	-10.5	12.5	0.401
Age	0.2	0.2	0.413
Education (<Bachelor degree=0)	1.9	5.6	0.731
Gender (Male =0)	3.3	5.3	0.540
Two approaches (Two mail surveys =0)	2.6	5.6	0.644
Mental Health Difference	R^2-adj =0		
Intercept	-2.0	4.2	0.639
Age	0.0	0.1	0.677
Education (<Bachelor degree=0)	0.6	1.9	0.750
Gender (Male =0)	0.5	1.8	0.795
Two approaches (Two mail surveys =0)	1.0	1.9	0.612
Vitality Difference	R^2-adj =0		
Intercept	1.0	4.2	0.816
Age	-0.1	0.1	0.451
Education (<Bachelor degree=0)	0.7	1.9	0.701
Gender (Male =0)	1.0	1.8	0.555
Two approaches (Two mail surveys =0)	2.1	1.9	0.260

Variable	β	SE	p Value
Bodily Pain Difference	R^2-adj = 0		
Intercept	6.7	5.4	0.222
Age	-0.1	0.1	0.249
Education (<Bachelor degree=0)	-1.7	2.5	0.495
Gender (Male =0)	3.6	2.3	0.121
Two approaches (Two mail surveys =0)	0.4	2.5	0.869
General Health Perception Difference	R^2-adj =0		
Intercept	1.0	4.7	0.839
Age	0.0	0.1	0.887
Education (<Bachelor degree=0)	2.9	2.1	0.183
Gender (Male =0)	-2.6	2.0	0.198
Two approaches (Two mail surveys =0)	-2.3	2.2	0.293
Physical Component Summary Difference	R^2-adj =0		
Intercept	0.5	2.6	0.855
Age	0.0	0.0	0.487
Education (<Bachelor degree=0)	0.4	1.2	0.713
Gender (Male =0)	1.3	1.2	0.260
Two approaches (Two mail surveys =0)	0.1	1.2	0.915
Mental Component Summary Difference	R^2-adj =0		
Intercept	-0.9	2.5	0.716
Age	0.0	0.0	0.760
Education (<Bachelor degree=0)	1.2	1.1	0.302
Gender (Male =0)	0.0	1.1	0.956
Two approaches (Two mail surveys =0)	0.3	1.1	0.790
WHOQOL-BREF			
Physical Difference	R^2-adj = 0		
Intercept	0.4	3.3	0.910
Age	0.0	0.1	0.779
Education (<Bachelor degree=0)	0.2	1.5	0.882
Gender (Male =0)	0.6	1.4	0.682
Two approaches (Two mail surveys =0)	-0.5	1.5	0.766
Psychological Difference	R^2-adj = 0		
Intercept	-7.5	3.3	0.023
Age	0.1	0.1	0.062
Education (<Bachelor degree=0)	2.4	1.5	0.099
Gender (Male =0)	1.9	1.4	0.175
Two approaches (Two mail surveys =0)	0.4	1.5	0.813

Variable	*β*	SE	*p* Value
Social (TW) [b] Difference	R^2-adj = 0		
Intercept	-0.8	3.3	0.820
Age	0.0	0.1	0.779
Education (<Bachelor degree=0)	1.6	1.5	0.280
Gender (Male =0)	-0.4	1.4	0.766
Two approaches (Two mail surveys =0)	-2.7	1.5	0.072
Environment (TW) [b] Difference	R^2-adj = 0		
Intercept	-0.9	2.8	0.753
Age	0.0	0.0	0.614
Education (<Bachelor degree=0)	-0.9	1.3	0.483
Gender (Male =0)	-0.8	1.2	0.501
Two approaches (Two mail surveys =0)	-0.2	1.3	0.867
EQ-5D			
Mobility Difference	R^2-adj = 0		
Intercept	0.0	0.0	0.953
Age	0.0	0.0	0.887
Education (<Bachelor degree=0)	0.0	0.0	0.678
Gender (Male =0)	0.0	0.0	0.025
Two approaches (Two mail surveys =0)	0.0	0.0	0.676
Self-care Difference	R^2-adj = 0		
Intercept	0.0	0.0	0.592
Age	0.0	0.0	0.335
Education (<Bachelor degree=0)	0.0	0.0	0.556
Gender (Male =0)	0.0	0.0	0.271
Two approaches (Two mail surveys =0)	0.0	0.0	0.656
Activities Difference	R^2-adj = 0		
Intercept	0.1	0.1	0.197
Age	0.0	0.0	0.242
Education (<Bachelor degree=0)	0.0	0.0	0.249
Gender (Male =0)	0.0	0.0	0.447
Two approaches (Two mail surveys =0)	0.0	0.0	0.624
Pain/discomfort Difference	R^2-adj = 0		
Intercept	0.0	0.1	0.771
Age	0.0	0.0	0.834
Education (<Bachelor degree=0)	0.0	0.1	0.958
Gender (Male =0)	-0.1	0.1	0.334
Two approaches (Two mail surveys =0)	0.1	0.1	0.056

Variable	β	SE	p Value
Anxiety/depression Difference	R^2-adj =0		
Intercept	0.1	0.2	0.468
Age	0.0	0.0	0.604
Education (<Bachelor degree=0)	0.0	0.1	0.624
Gender (Male =0)	0.0	0.1	0.870
Two approaches (Two mail surveys =0)	0.0	0.1	0.752
Utility Difference	R^2-adj =0		
Intercept	0.0	0.0	0.264
Age	0.0	0.0	0.331
Education (<Bachelor degree=0)	0.0	0.0	0.574
Gender (Male =0)	0.0	0.0	0.302
Two approaches (Two mail surveys =0)	0.0	0.0	0.237

* p<0.05

Table 5 Comparison of the floor and ceiling of the SF-36, WHOQOL-BREF, and EQ-5D of pre-test in hospital interview and mail surveys

SF-36				**WHOQOL -BREF**				**EQ-5D**			
	range of raw scores	**% at ceiling**	**% at floor**		**range of raw scores**	**% at ceiling**	**% at floor**		**range of raw scores**	**% at ceiling**	**% at floor**
Physical Functioning	9	34.6%	0.0%	Physical	14	1.9%	0.0%	Mobility	2	98.1%	0.0%
Social Functioning	7	25.0%	0.0%	Psychological	13	0.0%	0.0%	Self-care	1	100.0%	0.0%
Role Limitations Physical	5	57.7%	15.4%	Social (TW) [b]	10	0.0%	0.0%	Activities	2	94.2%	0.0%
Role Limitations Emotional	4	53.9%	21.2%	Environment (TW) [b]	22	0.0%	0.0%	Pain/ discomfort	2	80.8%	0.0%
Mental Health	14	3.9%	0.0%					Anxiety/ depression	2	75.0%	0.0%
Vitality	18	1.9%	0.0%								
Bodily Pain	10	36.5%	0.0%								
General health perception	24	0.0%	0.0%								

Table 6 Comparison of the floor and ceiling of the SF-36, WHOQOL-BREF, and EQ-5D of pre-test in two mail surveys

SF-36				**WHOQOL -BREF**				**EQ-5D**			
	range of raw scores	**% at ceiling**	**% at floor**		**range of raw scores**	**% at ceiling**	**% at floor**		**range of raw scores**	**% at ceiling**	**% at floor**
Physical Functioning	11	48.4%	0.0%	Physical	14	0.0%	0.0%	Mobility	2	97.6%	0.0%
Social Functioning	5	24.8%	0.0%	Psychological	18	0.0%	0.0%	Self-care	2	99.2%	0.0%
Role Limitations Physical	5	73.0%	8.7%	Social (TW)[b]	14	0.0%	0.0%	Activities	2	97.6%	0.0%
Role Limitations Emotional	4	70.4%	11.2%	Environment (TW)[b]	19	0.8%	0.0%	Pain/ discomfort	2	78.4%	0.0%
Mental Health	22	0.8%	0.0%					Anxiety/ depression	3	79.4%	1.6%
Vitality	18	0.8%	0.8%								
Bodily Pain	13	49.6%	0.0%								
General health perception	31	0.0%	0.0%								