<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Psychometrics of the chronic liver disease questionnaire for Southern Chinese patients with chronic hepatitis B virus infection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lam, ETP; Lam, CLK; Lai, CL; Yuen, MF; Fong, DYT</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>World Journal Of Gastroenterology, 2009, v. 15 n. 26, p. 3288-3297</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2009</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/59341">http://hdl.handle.net/10722/59341</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
BRIEF ARTICLES

Psychometrics of the chronic liver disease questionnaire for Southern Chinese patients with chronic hepatitis B virus infection

Elegance Ting Pui Lam, Cindy Lo Kuen Lam, Ching Lung Lai, Man Fung Yuen, Daniel Yee Tak Fong

Abstract

AIM: To test the psychometric properties of a Chinese ([Hong Kong] HK) translation of the chronic liver disease questionnaire (CLDQ).

METHODS: A Chinese (HK) translation of the CLDQ was developed by iterative translation and cognitive debriefing. It was then administered to 72 uncomplicated and 78 complicated chronic hepatitis B (CHB) patients in Hong Kong together with a structured questionnaire on service utilization, and the Chinese (HK) SF-36 Health Survey Version 2 (SF-36v2).

RESULTS: Scaling success was ≥ 80% for all but three items. A new factor assessing sleep was found and items of two (Fatigue and Systemic Symptoms) subscales tended to load on the same factor. Internal consistency and test-retest reliabilities ranged from 0.58-0.90 for different subscales. Construct validity was confirmed by the expected correlations between the SF-36v2 Health Survey and CLDQ scores. Mean scores of CLDQ were significantly lower in complicated compared with uncomplicated CHB, supporting sensitivity in detecting differences between groups.

CONCLUSION: The Chinese (HK) CLDQ is valid, reliable and sensitive for patients with CHB. Some modifications to the scaling structure might further improve its psychometric properties.

© 2009 The WJG Press and Baishideng. All rights reserved.

Key words: Chronic liver disease; Health-related quality of life; Hepatitis B; Southern Chinese; Validity

Peer reviewer: Edmund J Bini, Professor, VA New York Harbor Healthcare System, Division of Gastroenterology (111D), 423 East 23rd Street, New York 10010, United States


INTRODUCTION

Chronic hepatitis B (CHB) virus infection remains a major global health problem. It is estimated that 350 million people worldwide are chronically infected, of whom one third (120 million) are Chinese[1]. The prevalence is higher in southern China (> 10%) than Northern China (6%-10%)[2]. Up to 25% of patients may die from CHB complications, such as cirrhosis-related complications or hepatocellular carcinoma (HCC), posing a threat to both mental and physical health, leading to impairment of health-related quality of life (HRQOL).

HRQOL has become an important outcome measure in clinical and health policy settings in the last two decades. Disease-specific measures are often needed to complement generic measures to give a more comprehensive evaluation of the HRQOL of patients with specific diseases. Several HRQOL measures have been developed specifically for chronic liver disease (CLD), such as the Chronic Liver Disease Questionnaire (CLDQ)[3], the Hepatitis Quality of Life (HQLQ)[4], the Liver Disease Quality of Life[5] and the Liver Disease Symptom Index (LDSI)[6]. The CLDQ developed by Younossi et al[7] was the first and is the most widely used. The other liver disease-specific HRQOL measures are not commonly used because they are either too long, or the validity data are limited[4,8].
The CLDQ consists of 29 items which are grouped into 6 subscales: abdominal symptoms (AS), fatigue (FA), systemic symptoms (SS), activity (AC), emotional function (EF) and worry (WO). It is applicable to all types of liver diseases including CHB. It has been shown to have adequate internal reliability, validity and sensitivity. Test-retest reliability was more variable with intra-class correlation (ICC) ranging from 0.23 to 0.72 for different subscales[3]. Previous studies showed that the CLDQ is more responsive than a generic measure to detect a change in patients with CLD[8]. It has been translated and validated in different languages[9-13], supporting its potential for cross-cultural adaptation. However, most of the psychometric data of the CLDQ have been derived from patients with hepatitis C virus (HCV) infection and Western populations. There are few data on its applicability for Southern Chinese CHB patients despite the fact that China has the world’s largest population suffering from CLD.

Recently, the CLDQ has been translated into Mandarin Chinese but this Chinese (Mainland) version may not be applicable to Southern Chinese who speak Cantonese, a dialect that has significant differences in the usage of words and terms from Mandarin. In addition, information on the validity, reliability and other psychometric properties of the Chinese (Mainland) CLDQ version is limited. The aim of this study was to test the psychometric properties of a Chinese (Hong Kong [HK]) translation of the CLDQ for Southern Chinese CHB patients. This would enable the evaluation of the impact of CHB infection and assess the effect of anti-viral drug treatments on HRQOL in the world’s largest population of CHB patients.

The objectives of this study were: (1) To develop a Chinese (HK) CLDQ that is semantically equivalent to the original; (2) To test the scaling assumptions and factor structure of the Chinese (HK) CLDQ; (3) To assess the psychometric properties in terms of reliability, construct validity, and sensitivity of the Chinese (HK) CLDQ; (4) To determine whether any modification of the CLDQ can improve its psychometric properties for Southern Chinese CHB patients.

MATERIALS AND METHODS

Ethics
This research project was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (IRB reference No., UW 06-089 T/1114 and trial registration No., HKCTR-151).

Development of the Chinese (HK) CLDQ
The Chinese (HK) translation of the CLDQ was developed by iterative translations, expert panel review and cognitive debriefing, as recommended guidelines by experts[14-15]. The original CLDQ was translated into Chinese by two independent professional translators. Reconciliation of the forward translations into a single forward translation was carried out by a bilingual expert in HRQOL measures (Lam CLK) and the translators. The reconciled Chinese translation was back-translated into English by another professional translator. The back translation was reviewed by the original author and the bilingual expert to identify any non-equivalence in the Chinese translation, which was then revised. The first draft of the Chinese (HK) CLDQ was evaluated by cognitive debriefing interviews with six Southern Chinese patients with CHB infection and further revision was made to ensure item clarity and equivalence to become the final Chinese (HK) CLDQ (used in this study on psychometrics properties).

Subjects
Patients with complicated CHB were recruited from outpatient hepatitis clinics of a regional hospital and patients with uncomplicated CHB were randomly selected from the computerized registers of three public primary care clinics serving over 100,000 people in one of five regions in Hong Kong. Patients aged 18 years or older who were hepatitis B surface antigen-positive for more than six months were included in the study. Patients were excluded if they could not communicate in Cantonese; had cognitive impairment shown by the patient’s inability to understand the study to give consent; were co-infected with HIV, HCV or hepatitis D virus; had undergone liver transplantation or had end-stage non-hepatitis B-related illnesses; were currently taking excessive alcohol (> 30 U/wk) or illegal drugs; or refused to give consent. Each patient completed the Chinese (HK) CLDQ, the Chinese (HK) SF-36v2 Health Survey and a structured questionnaire on morbidity and socio-demographics, administered by a trained interviewer. Each patient was asked if he/she had ever been diagnosed by a registered practitioner for more than four weeks to have hypertension, diabetes mellitus, heart disease, stroke, chronic lung disease, arthritis, psychological illness (i.e. depression, anxiety, neurasthenia or psychosis) or any other chronic diseases. Chronic co-morbidity was measured by the total number of diseases (summation of positive responses to the questions) and the presence of a specific diagnosis. Clinical data related to the CHB infection including Child’s staging for patients with cirrhosis and the biomarkers of liver disease (alanine aminotransferase, aspartate aminotransferase, α-fetoprotein and total bilirubin) in each patient were retrieved from medical records. Socio-demographic data including age, gender, education, marital status, occupation, household income and family history of liver disease were also collected.

The Chinese (HK) CLDQ was re-administered to the 46 subjects with uncomplicated CHB, whose condition was expected to be stable, by telephone two weeks from the first administration, in order to assess the test-retest reliability of the Chinese (HK) CLDQ. Sixty one percent of the repeat interviews were carried out by the same interviewer.

Instruments
The Chinese (HK) CLDQ consists of 29 items measuring six subscales as described above. Each item is rated on a 7-point (1 = all of the time to 7 = none of the time) Likert scale. Scores for each of the six domains are calculated by the mean of the item scores within the subscale. A summary score is calculated by the mean of all subscale scores. The scores range from 1 to 7 with a higher score indicating better HRQOL.
The Chinese (HK) SF-36v2 Health Survey is a generic HRQOL measure that has been translated, validated and normed on the general Chinese population in Hong Kong\[16,17\]. It measures eight domains of HRQOL on physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-motional (RE) and mental health (MH). Summations of item scores of the same domain give the domain scores, which are transformed into a range from 0 to 100. A higher score indicates better HRQOL. The eight domain scores are summarized to form the physical component (PCS) and mental component (MCS) summary scores.

**Statistical analysis**
All data analysis was carried out in SPSS for Windows 15.0. Statistical significant levels were set at \( P \) values less than 0.05.

**Scaling assumptions**
The CLDQ item and subscale scores were calculated and tested against the following scaling assumptions: (1) Items should be substantially linearly correlated to the hypothesized subscale score with a coefficient of 0.4 or above by Spearman rank correlation test, to show the item is a significant indicator for the subscale concept. (2) An item should have a stronger correlation with its hypothesized subscale than other subscales indicating scaling success\[18\]. This is a test of item discriminant validity. The difference between correlations is statistically significant if it is greater than two standard errors (1 divided by the square root of sample size).

**Construct validity**

**Factor analysis:** Exploratory factor analysis using principal components with varimax rotation was performed to evaluate the factor structure of the Chinese (HK) CLDQ. The criterion for factor extraction was an eigenvalue greater than one. The highest factor loading was chosen for each item. The scree plot was also used to determine the number of factors.

**Convergent validity:** Construct validity was also tested by convergent validity determined by Spearman correlations between corresponding CLDQ and SF-36v2 Health Survey domain scores. It was hypothesized that moderate \( (r = 0.4 \text{ to } 0.7) \) to strong \( (r > 0.7) \) correlations should exist between CLDQ FA and SF-36v2 VT; between CLDQ SS and SF-36v2 BP; between CLDQ AC and SF-36v2 PF; RP and RE; and between CLDQ EF and SF-36v2 MH scores.

**Sensitivity**
The mean CLDQ scores were compared between two CHB patient groups, and the significance was tested by independent \( t \) to evaluate its sensitivity in detecting a difference between patients with complicated and uncomplicated infections. The sensitivity of the CLDQ was also assessed by the effect size (difference between group mean scores/overall standard deviation). According to Cohen\[19\], effect sizes of 0.3, 0.5 and 0.8 were considered small, medium and large differences, respectively. An effect size of less than 0.3 was considered not significant.

**Re-liability**
Different methods were used to assess reliability, including internal consistency and test-retest reliability. Internal consistency was measured by Cronbach’s \( \alpha \), which is a measure of the extent to which items in a questionnaire are homogeneous (correlated) in supporting the same concept\[20\]. Test-retest reliability refers to the stability of an instrument over time\[21\], which was measured by the intra-class correlation (ICC) between the two-week test-retest results. Reliability coefficients \( \geq 0.7 \) and 0.9 are usually expected for group comparisons and individual comparisons, respectively\[22\].

**Results**

**Translational equivalence of the Chinese (HK) CLDQ**
All items except item 11 (level of energy) were found to be understood by 6 patients. Three out of six patients did not understand item 11. Five patients (83%) interpreted the meaning of all except four items (11, 13, 19 and 28) correctly. Four out of six patients misinterpreted the meaning of item 11 with three interpreting it as decreased physical strength. Two patients (33.3%) had difficulty in differentiating the meaning of “sleepy” and “drowsy”; and did not seem to have interpreted the words “mood swings” (item 19). Two out of six patients did not include the meaning of “worried about never feeling better” (item 28) in their interpretation. The Chinese (HK) translation was revised based on the results of cognitive debriefing and the revised questionnaire was then field tested on 23 CHB patients before this study. The final Chinese (HK) CLDQ was formed and its back-translation is shown in the appendix.

**Subjects**
One hundred and eighty four CHB patients were identified; 6 patients were excluded (3 had hepatitis B infection less than 6 mo, 2 had communication problems and 1 had co-infection with HCV) and 28 patients refused to participate in this study. One hundred and fifty Chinese adults consisting of 72 uncomplicated (normal liver function defined as liver enzymes persistently within the normal range and without any history of cirrhosis or HCC) and 78 complicated (cirrhosis or HCC) completed the study, giving a response rate of 84.3% (150/178). Table 1 shows their characteristics, overall and by disease severity groups. There were 8 patients in the complicated CHB group who had HCC without any cirrhosis and had normal liver function. There were no statistical differences in demographics between the uncomplicated and complicated CHB groups, except age and sex \( (P < 0.001) \). Complicated CHB patients were older and there were more men than those in the uncomplicated group which was expected because CHB complications were more common in men than in women and the median age for the development of complications was 57.2 years\[22,23\].

**Score distribution**
Table 2 shows the distribution of the Chinese (HK) CLDQ and SF-36v2 scores. There was practically no
floor effect but there were significant ceiling effects in the Chinese (HK) CLDQ AS, AC and WO subscales, more so in the uncomplicated than the complicated group. Significant ceiling effects were also found in most SF-36v2 Health Survey scales. Sub-group analysis showed that the mean Chinese (HK) CLDQ scores were significantly lower in the complicated group than the uncomplicated group in all subscale and overall scores.

Figure 1 compares the distribution of the Chinese (HK) CLDQ scores with those from other countries. The distribution pattern of the Chinese (HK) CLDQ subscale scores was very similar to those of other countries except Italy, supporting cross-cultural conceptual equivalence.

Scaling properties
Table 3 shows the mean item scores and standard deviation of the 29 CLDQ items grouped under their hypothesized subscales. All correlations between items and their hypothesized subscale score exceeded the standard of 0.4. All but six items had a higher correlation with its hypothesized subscale than other subscales, i.e. 100% scaling success. Four items of the SS subscale and two items of the AC subscale correlated more highly with some other subscales than their own. Scaling success was the lowest in item 3 “bodily pain”, which correlated more highly with four other subscales than with the SS subscale, with the highest found for EF, but the differences were not statistically significant. Items 6 “shortness of breath”, 23 “dry mouth”, 27 “itching”, 7 “not able to eat as much as you would like” and 9 “trouble in lifting or carrying heavy objects” correlated higher with one to three other subscales than its own, but the differences in the correlations were not statistically significant.

The overall scaling success rate on discriminant validity was 100% for four scales (AS, FA, EF and WO), but it was 73% for the AC subscale and 64% for the SS subscale.

Factor analysis
Table 4 illustrates the rotated factor loadings between the
29 items and 6 factors with eigenvalue > 1. The six factors explained 70.1% of total variance. The factor loadings of the items were not entirely consistent with the scaling hypothesis. Items of the FA and SS subscales, except bodily pain (item 3), decreased strength (item 8) and scales assessing sleep (items 16 and 20). The factors of EF, WQ, AS and AC subscales loaded nicely on their hypothesized factors.

**Construct validity**

Table 5 shows the correlations between the scores of the CLDQ and SF-36v2 Health Survey. As hypothesized, moderate to strong correlations were found between CLDQ FA and SF-36v2 VT scores; and between CLDQ SS and the SF-36v2 BP scores. The CLDQ AC score correlated significantly with all SF-36v2 Health Survey domain scores and the strongest was found with the SF-36v2 BP score. The CLDQ EF score correlated strongly not only with the SF-36v2 MH score but moderately with the SF-36v2 VT, RE, RP and GH scores.

**Sensitivity**

As shown in Table 2, the CLDQ overall and subscale mean scores were all significantly higher in the uncomplicated than the complicated CHB group. The effect sizes of the group differences in the CLDQ scores all exceeded 0.4 (range 0.4-0.6). Only three of the eight SF-36v2 domain scores (PF, RP and SF) and the PCS score detected a significant difference between the uncomplicated and the complicated groups. However, the greatest effect size difference between the two groups was found in the SF-36v2 RP score.

**Reliability**

Across all subscales, the Cronbach’s $\alpha$ coefficients of the internal consistency reliability were higher than the recommended value of 0.7 (Table 5). ICC coefficients measuring the two-week test-retest reliability exceeded 0.7 in all but the AS (0.58) and AC (0.66) subscales. The reliability coefficients were comparable to those of the SF-36v2 Health Survey.

Table 5 also shows that the correlations (range 0.50-0.87) between the CLDQ subscales were smaller than the subscale internal reliability coefficients for all subscales, showing that each subscale measures a distinct concept. The overall CLDQ scores correlated strongly with all CLDQ subscale scores.

**DISCUSSION**

The mean scores of the CLDQ found in our population were generally higher than those found in other countries. This might be the result of a sampling difference, in that over half of our subjects had uncomplicated CHB infection and most of the other studies included patients with more serious diseases and patients with HCV who tend to have more impairment in HRQOL than patients with CHB infection. The other reason for a difference in the absolute HRQOL scores between different populations is a difference in the sociocultural norms. A comparison
with the population norms of generic HRQOL measures such as those of the SF-36 Health Survey will provide a more meaningful interpretation on the impact of CHB on HRQOL between different populations. Our study found that uncomplicated CHB patients had significant impairment in the SF-36v2 RP, BP, SF and RE domains, and complicated CHB patients had significantly lower SF-36v2 scores in six domains (PF, RP, BP, GH, SF and RE) than the norms of the HK population (Table 2)\(^{25,26}\). The findings suggested that CHB infection affected HRQOL only modestly unless complications develop. Surprisingly, there was no difference in the MH score between CHB patients and the HK population norm. It was unlikely that a potentially lethal chronic infection had no effect on mental health, the SF-36v2 Health Survey was probably not sensitive enough to detect the difference.

The high ceiling effects in the AS, AC and WO subscales in patients with uncomplicated CHB were expected since they were usually asymptomatic. A pattern that was similar to that found in a Spanish population\(^{9}\). A high ceiling effect was also observed among patients with complicated CHB which was unexpected, this was probably because most of our subjects with complicated CHB were under anti-viral treatment that might have improved their HRQOL, or perhaps some patients had adjusted to their illnesses. On the whole, the CLDQ had a lower ceiling effect than the SF-36v2 Health Survey, suggesting that this disease-specific HRQOL measure would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies. The lack of floor effect indicates that the Chinese (HK) CLDQ would be more responsive than the generic measure in detecting improvements with treatment, which needs to be confirmed by prospective studies.
CLDQ in general. However, the scaling success rates of items 3 (bodily pain), 23 (dry mouth) and 9 (trouble in lifting or carrying heavy objects) seemed too low to be acceptable, raising the question whether they should be

Table 4  Factor loadings of the Chinese (HK) CLDQ

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1 EF</th>
<th>Factor 2 WO</th>
<th>Factor 3 SS + FA</th>
<th>Factor 4 AS</th>
<th>Factor 5 AC</th>
<th>Factor 6 SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal symptom (AS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: Abdominal bloating</td>
<td>0.18</td>
<td>0.27</td>
<td>0.16</td>
<td>0.70</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>5: Abdominal pain</td>
<td>0.20</td>
<td>0.06</td>
<td>0.13</td>
<td>0.81</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>17: Abdominal discomfort</td>
<td>0.17</td>
<td>0.23</td>
<td>0.17</td>
<td>0.82</td>
<td>0.20</td>
<td>0.08</td>
</tr>
<tr>
<td>Fatigue (FA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2: Tiredness or fatigue</td>
<td>0.33</td>
<td>0.03</td>
<td>0.51</td>
<td>0.31</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>4: Feel sleepy during the day</td>
<td>0.28</td>
<td>-0.05</td>
<td>0.47</td>
<td>0.14</td>
<td>0.44</td>
<td>0.14</td>
</tr>
<tr>
<td>8: Decreased strength</td>
<td>0.37</td>
<td>0.14</td>
<td>0.37</td>
<td>0.24</td>
<td>0.60</td>
<td>0.24</td>
</tr>
<tr>
<td>11: Decreased energy</td>
<td>0.56</td>
<td>0.06</td>
<td>0.37</td>
<td>0.11</td>
<td>0.51</td>
<td>0.12</td>
</tr>
<tr>
<td>13: Drowsiness</td>
<td>0.49</td>
<td>0.09</td>
<td>0.47</td>
<td>0.15</td>
<td>0.36</td>
<td>-0.17</td>
</tr>
<tr>
<td>Systemic symptoms (SS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Bodily pain</td>
<td>0.49</td>
<td>0.00</td>
<td>0.18</td>
<td>0.39</td>
<td>0.14</td>
<td>0.22</td>
</tr>
<tr>
<td>6: Shortness of breath</td>
<td>0.22</td>
<td>0.22</td>
<td>0.59</td>
<td>0.39</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>21: Muscle cramps</td>
<td>0.09</td>
<td>0.08</td>
<td>0.76</td>
<td>0.17</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>23: Dry mouth</td>
<td>0.24</td>
<td>0.42</td>
<td>0.48</td>
<td>0.16</td>
<td>0.08</td>
<td>-0.02</td>
</tr>
<tr>
<td>27: Itching</td>
<td>0.10</td>
<td>0.27</td>
<td>0.65</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.28</td>
</tr>
<tr>
<td>Activity (AC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7: Not able to eat as much as you would like</td>
<td>0.05</td>
<td>0.29</td>
<td>0.04</td>
<td>0.24</td>
<td>0.81</td>
<td>0.12</td>
</tr>
<tr>
<td>9: Trouble in lifting or carrying heavy objects</td>
<td>0.41</td>
<td>0.18</td>
<td>0.37</td>
<td>-0.05</td>
<td>0.35</td>
<td>0.17</td>
</tr>
<tr>
<td>14: Bothered by a limitation of the diet</td>
<td>0.11</td>
<td>0.43</td>
<td>0.09</td>
<td>0.10</td>
<td>0.67</td>
<td>0.12</td>
</tr>
<tr>
<td>Emotional function (EF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10: Anxiety</td>
<td>0.67</td>
<td>0.32</td>
<td>0.15</td>
<td>0.39</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>12: Unhappiness</td>
<td>0.77</td>
<td>0.25</td>
<td>0.15</td>
<td>0.20</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>15: Irritability</td>
<td>0.78</td>
<td>0.27</td>
<td>0.15</td>
<td>0.17</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>16: Difficulty in sleeping at night</td>
<td>0.27</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.14</td>
<td>0.79</td>
</tr>
<tr>
<td>19: Mood swings</td>
<td>0.78</td>
<td>0.24</td>
<td>0.10</td>
<td>0.19</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>20: Difficulty in falling asleep at night</td>
<td>0.16</td>
<td>0.14</td>
<td>0.25</td>
<td>0.22</td>
<td>0.25</td>
<td>0.71</td>
</tr>
<tr>
<td>24: Depression</td>
<td>0.80</td>
<td>0.31</td>
<td>0.15</td>
<td>0.08</td>
<td>0.03</td>
<td>0.22</td>
</tr>
<tr>
<td>26: Problems with concentration</td>
<td>0.52</td>
<td>0.28</td>
<td>0.29</td>
<td>0.18</td>
<td>0.37</td>
<td>0.27</td>
</tr>
<tr>
<td>Worry (WO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18: Worries about the impact of the liver disease</td>
<td>0.40</td>
<td>0.47</td>
<td>0.03</td>
<td>0.50</td>
<td>0.25</td>
<td>0.07</td>
</tr>
<tr>
<td>22: Worries that symptoms will develop into major problem</td>
<td>0.29</td>
<td>0.80</td>
<td>0.11</td>
<td>0.16</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>25: Worries that the condition is getting worse</td>
<td>0.29</td>
<td>0.84</td>
<td>0.13</td>
<td>0.17</td>
<td>0.17</td>
<td>0.09</td>
</tr>
<tr>
<td>28: Worries about never feeling any better</td>
<td>0.27</td>
<td>0.77</td>
<td>0.13</td>
<td>0.18</td>
<td>0.26</td>
<td>0.24</td>
</tr>
<tr>
<td>29: Availability of a liver for transplant</td>
<td>0.16</td>
<td>0.64</td>
<td>0.26</td>
<td>0.14</td>
<td>0.30</td>
<td>0.10</td>
</tr>
</tbody>
</table>

SL: Sleep.

Table 5  Reliability and correlations of the Chinese (HK) CLDQ and SF-36v2 health survey scores

<table>
<thead>
<tr>
<th>CLDQ</th>
<th>AS</th>
<th>FA</th>
<th>SS</th>
<th>AC</th>
<th>EF</th>
<th>WO</th>
<th>Cronbach’s α</th>
<th>ICC 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>0.84</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>0.54</td>
<td>0.88</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>0.54</td>
<td>0.72</td>
<td>0.74</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.50</td>
<td>0.67</td>
<td>0.60</td>
<td>0.72</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF</td>
<td>0.58</td>
<td>0.72</td>
<td>0.68</td>
<td>0.61</td>
<td>0.90</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WO</td>
<td>0.52</td>
<td>0.57</td>
<td>0.60</td>
<td>0.58</td>
<td>0.69</td>
<td>0.90</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.72</td>
<td>0.87</td>
<td>0.84</td>
<td>0.80</td>
<td>0.86</td>
<td>0.81</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>SF-36v2 Health Survey</td>
<td>0.47</td>
<td>0.56</td>
<td>0.65</td>
<td>0.54</td>
<td>0.47</td>
<td>0.46</td>
<td>0.81</td>
<td>0.93</td>
</tr>
<tr>
<td>PF</td>
<td>0.56</td>
<td>0.72</td>
<td>0.70</td>
<td>0.67</td>
<td>0.60</td>
<td>0.62</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>RP</td>
<td>0.44</td>
<td>0.44</td>
<td>0.62</td>
<td>0.42</td>
<td>0.48</td>
<td>0.44</td>
<td>0.89</td>
<td>0.77</td>
</tr>
<tr>
<td>BP</td>
<td>0.46</td>
<td>0.66</td>
<td>0.56</td>
<td>0.57</td>
<td>0.60</td>
<td>0.62</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>GH</td>
<td>0.51</td>
<td>0.79</td>
<td>0.63</td>
<td>0.58</td>
<td>0.67</td>
<td>0.54</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>VT</td>
<td>0.44</td>
<td>0.56</td>
<td>0.52</td>
<td>0.60</td>
<td>0.58</td>
<td>0.52</td>
<td>0.88</td>
<td>0.54</td>
</tr>
<tr>
<td>RE</td>
<td>0.53</td>
<td>0.51</td>
<td>0.52</td>
<td>0.46</td>
<td>0.66</td>
<td>0.50</td>
<td>0.89</td>
<td>0.74</td>
</tr>
<tr>
<td>MH</td>
<td>0.41</td>
<td>0.58</td>
<td>0.49</td>
<td>0.49</td>
<td>0.78</td>
<td>0.61</td>
<td>0.84</td>
<td>0.89</td>
</tr>
</tbody>
</table>

1 All analyses were performed with the total sample of 150 patients, except for the Intra-class Correlation Coefficient (ICC), which was based on the results from 46 uncomplicated CHB patients.
grouped under other subscales than the originally hypothesized. It is interesting to note that bodily pain correlated the most with the EF subscale score and loaded the strongest on the EF factor (Table 4). It is a common observation that emotional state has a strong influence on pain perception and vice versa. Although the items on dry mouth or trouble in lifting or carrying heavy objects correlated more strongly with other subscales than their own, they should probably remain in the hypothesized subscale because the differences in the item-subscale correlations were not significant and the item-hypothesized subscale correlations were greater than 0.4. Furthermore, the factor loading results were not consistent with the results of the item-subscale correlations. The item “dry mouth” correlated most strongly with WO subscale score but the loading was the highest on the SS factor (0.48). The item “trouble in lifting or carrying heavy objects” correlated highest with the SS score but factor analysis showed that it loaded most strongly on the factor of EF.

The factor structure of the Chinese (HK) CLDQ version was almost identical to the original CLDQ in four subscales (EF, WO, AS and AC). The new factor of Sleep found in our Chinese population was also found in the Spanish, Italian and German populations. CHB patients may have sleep difficulties due to reasons other than emotional problems, such as pain and other symptoms. The items of the FA and SS subscales, except items 3, 8, and 11, loaded on one single factor since they all measure symptoms. Items 3 and 11 of the FA subscale loaded on the AC factor. Factor analysis with promax rotation was also performed to cross-validate the factor structure obtained by the varimax rotation, and it showed similar results with a new factor assessing Sleep and items of the FA subscale loaded mostly on the AC factor instead of a separate factor.

An alternative scaling structure for the Chinese (HK) CLDQ based on the factor loading results could be formed. Items 16 (difficulty in sleeping) and 20 (difficulty in falling asleep at night) were grouped into a new Sleep subscale. Items 8 (decreased strength) and 11 (decreased energy) were grouped into the AC subscale. Items 2, 4 and 13 of the original FA subscale are grouped with items of the SS subscale to form the new SS subscale. Item 3 (bodily pain), although loaded most strongly and correlated the most with EF factor, remains in the SS subscale because this has better face validity. The psychometric properties of the revised Chinese (HK) CLDQ subscales with regrouping of the items are shown in Table 6. It can be seen that the new subscale structure greatly improves the scaling success rates of the SS and AC items, although it reduces the success rate of the EF subscale slightly. The new scaling structure also reduced the ceiling effects of the SS and AC subscales. Further studies are needed to determine whether the revised subscale structure will translate into better sensitivity and responsiveness in clinical applications. Until such data are available, the original subscale structure of the CLDQ is recommended to allow better international comparability.

The expected correlations between the CLDQ and SF-36v2 Health Survey domains were observed confirming convergent construct validity. The correlation with the SF-36v2 RE domain was higher in the CLDQ EF than the AC subscale because conceptually the SF-36v2 RE measures the effect of emotional problems on daily activities.

The CLDQ subscales of AS and WO address domains that are not assessed by the generic measure (SF-36v2 Health Survey) and detected significant differences between the two groups of CHB patients. There were significant differences in the WO and EF subscales of the CLDQ between the CHB groups although this was not found in most of the mental-health related domains (RE, MH and MCS) of the SF-36v2 Health Survey, suggesting that the Chinese (HK) CLDQ was more sensitive than the generic measure in detecting the emotional impact of CHB. It is worth noting that although more domains in the CLDQ showed a significant difference between the complicated and uncomplicated CHB groups, the largest effect size difference was found in the SF-36v2 RP domain indicating that a disease-specific measure may not always be more sensitive than a generic measure. The two types of HRQOL measures should complement each other in the evaluation of the HRQOL of CHB patients.

Internal consistency and test-retest reliability were acceptable for all subscales. Test-retest reliability (ICC) of the AS subscale was relatively low (0.58) probably because these symptoms could fluctuate from day to day and pain intensity might vary noticeably in a relatively short period of time. Reliability of the CLDQ in our study was generally higher than those found in other studies (0.46-0.95). The SS subscale had very good test-retest reliability (ICC 0.86) in our population. The very low ICC (0.25) found in the US study was likely the result of an inappropriately long retest interval of six months.

Our study administered the Chinese (HK) CLDQ using an interviewer since our populations had a relatively low literacy level. The performance of the instrument by self-completion will need to be confirmed by further studies. The responsiveness of the Chinese (HK) CLDQ in detecting changes with disease progression or anti-viral treatment will also need to be determined.

---

Table 6 Psychometric properties of the revised Chinese (HK) CLDQ subscales

<table>
<thead>
<tr>
<th>Revised subscales</th>
<th>Mean</th>
<th>SD</th>
<th>% floor</th>
<th>% ceiling</th>
<th>Scaling success</th>
<th>Cronbach’s α</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>5.9</td>
<td>1.3</td>
<td>0.0</td>
<td>37.3</td>
<td>100.0</td>
<td>0.84</td>
<td>0.58</td>
</tr>
<tr>
<td>SS + FA</td>
<td>5.2</td>
<td>1.1</td>
<td>0.0</td>
<td>4.7</td>
<td>95.0</td>
<td>0.84</td>
<td>0.88</td>
</tr>
<tr>
<td>AC</td>
<td>5.6</td>
<td>1.4</td>
<td>0.0</td>
<td>20.7</td>
<td>88.0</td>
<td>0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>EF</td>
<td>5.3</td>
<td>1.3</td>
<td>0.0</td>
<td>12.0</td>
<td>93.3</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td>WO</td>
<td>5.4</td>
<td>1.6</td>
<td>0.7</td>
<td>27.3</td>
<td>100.0</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>SL</td>
<td>4.8</td>
<td>1.7</td>
<td>1.3</td>
<td>17.3</td>
<td>100.0</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>Overall Scores</td>
<td>5.4</td>
<td>1.1</td>
<td>0.0</td>
<td>0.7</td>
<td>NA</td>
<td>0.89</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Appendix: Back-translation of Chinese (HK) CLDQ

Original Wording

This questionnaire is designed to find out how you have been feeling during the last 2 wk. You will be asked about your symptoms related to your liver disease, how you have been affected in doing activities, and how your mood has been.

Please complete all of questions and select only one response for each question.

1 How much of the time during the last 2 wk have you been troubled by a feeling of abdominal bloating?

All of the time
Most of the time
A good bit of the time
A little of the time
Hardly any of the time
None of the time

2 How much of the time have you been tired or fatigued during the last 2 wk?

3 How much of the time during the last 2 wk have you experienced bodily pain?

4 How often during the last 2 wk have you felt sleepy during the day?

5 How much of the time during the last 2 wk have you experienced abdominal pain?

6 How much of the time during the last 2 wk have shortness of breath been a problem for you in your daily activities?

7 How much of the time during the last 2 wk have you not been able to eat as much as you would like?

8 How much of the time in the last 2 wk have you been bothered by having decreased strength?

9 How often during the last 2 wk have you had trouble lifting or carrying heavy objects?

10 How often during the last 2 wk have you felt anxious?

11 How often during the last 2 wk have you felt a decreased level of energy?

Backward Translation

The purpose of this questionnaire is to understand how you felt in the past 2 wk. The questions are about the symptoms resulting from your liver illness and how these symptoms affect your participation in activities, and your emotions. Please answer all questions. You can only choose one answer for each question.

1 In the past 2 wk, how much time have you been bothered by your bloating problem?

All of the time
Most of the time
Quite Often
Sometimes
A Short Time
Hardly Any
Never

2 In the past 2 wk, how much time did you feel tired or exhausted?

3 In the past 2 wk, how much time did you experience body ache?

4 In the past 2 wk, how often did you feel sleepy during the daytime?

5 In the past 2 wk, how much time did you have abdominal pain?

6 In the past 2 wk, how much time were your daily activities affected by your shortness of breath?

7 In the past 2 wk, how much time were you unable to eat as much as you want?

8 In the past 2 wk, how much time have you been bothered by the decline in your physical energy?

9 In the past 2 wk, how often did you find it difficult when you were lifting or carrying heavy objects?

10 In the past 2 wk, how often did you feel anxious?

11 In the past 2 wk, how often did you find your energy level decreasing?

12 How much of the time during the last 2 wk have you felt unhappy?

13 How often during the last 2 wk have you felt drowsy?

14 How much of the time during the last 2 wk have you been bothered by a limitation of your diet?

15 How often during the last 2 wk have you been irritable?

16 How much of the time during the last 2 wk have you had difficulty sleeping at night?

17 How much of the time during the last 2 wk have you been troubled by a feeling of abdominal discomfort?

18 How much of the time during the last 2 wk have you been worried about the impact your liver disease has on your family?

19 How much of the time during the last 2 wk have you had mood swings?

20 How much of the time during the last 2 wk have you been unable to fall asleep at night?

21 How often during the last 2 wk have you had muscle cramps?

22 How much of the time during the last 2 wk have you been worried that your symptoms will develop into major problems?

23 How much of the time during the last 2 wk have you had a dry mouth?

24 How much of the time during the last 2 wk have you felt depressed?

25 How much of the time during the last 2 wk have you been worried about your condition getting worse?

26 How much of the time during the last 2 wk have you had problems concentration?

27 How much of the time have you been troubled by itching during the last 2 wk?

28 How much of the time during the last 2 wk have you been worried about never feeling any better?

29 How much of the time during the last 2 wk have you been concerned about the availability of a liver if you need a liver transplant?

30 How much of the time during the last 2 wk have you felt happy?

31 How much of the time during the last 2 wk have you been able to do the things you want to do?

32 How much of the time during the last 2 wk have you felt preoccupied?

33 How much of the time during the last 2 wk have you had a limitation of your diet?

34 How much of the time during the last 2 wk have you been worried about your condition deteriorating?

35 How much of the time during the last 2 wk have you been worried about the availability of a liver if you need a liver transplant?

36 How much of the time during the last 2 wk have you been bothered by your bloating problem?

37 How much of the time during the last 2 wk have you been worried about your condition deteriorating?

38 How much of the time during the last 2 wk have you been worried about a liver if you need a liver transplant?
**Innovations and breakthroughs**

The Chronic Liver Disease Questionnaire (CLDQ) had been applied mainly to patients with hepatitis C virus infection in Western countries. This study was the first to show that a Chinese (Hong Kong) translation of the CLDQ was valid, reliable and sensitive for Southern Chinese patients with CHB infection. The CLDQ can be applied to millions of Southern Chinese CHB patients to evaluate their HRQOL. Some modifications might further improve its validity, reliability and sensitivity.

**Applications**

The Chinese (Hong Kong) CLDQ can be used to evaluate the impact of CHB virus infection and assess the effectiveness of anti-viral drug treatments in Cantonese-speaking Southern Chinese. The CLDQ can be used as a cross-cultural HRQOL measure in international studies that include Southern Chinese.

**Terminology**

CHB virus infection refers to those who are hepatitis B surface antigen-positive for more than six months. Validity is defined as the extent to which a test measures what it is intended to measure. Reliability refers to the consistency or stability of the measurement process across time, patients or observers.

**Peer review**

The authors validated and tested the psychometric properties of a Southern Chinese translation of the CLDQ and determined that their questionnaire was valid, reliable, and sensitive for southern Chinese patients with hepatitis B virus infection. The study was well done and used appropriate methodology to validate and test the questionnaire.

**REFERENCES**

18. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959; 56: 81-105