Reliability and Validity of a Chinese Version’s Health-Related Quality of Life Questionnaire for Hepatitis B Patients

Siew Chin Ong, BPharm (Hons), PhD,1 Seng Gee Lim, MBBS (Hons), FRACP, FRCP, FAMS, MD,2,3,4* Shu Chuen Li, BPham, GradDipBus(Tech Mgt), MAppiSc, MBA, PhD5*

1Department of Pharmacy, National University of Singapore, Singapore; 2Department of Gastroenterology and Hepatology, National University Hospital, Singapore; 3Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 4Centre for Molecular Medicine, Agency for Science, Technology and Research, Biopolis, Singapore; 5Discipline of Pharmacy & Experimental Pharmacology, School of Biomedical Sciences, University of Newcastle, Callaghan, New South Wales, Australia

ABSTRACT

Objectives: To culturally adapt a Chinese version of the Hepatitis Quality of Life Questionnaire (HQLQ) and assess its suitability for use in Chinese-speaking hepatitis B virus (HBV) patients in Singapore.

Study: Reliability was assessed using Cronbach’s alpha coefficients and intra-class correlation coefficients. Item-to-scale correlation was assessed using Spearman’s rank correlations (p) between scale scores and their constituent items. Convergent and divergent construct validities were tested in three and two a priori hypotheses, respectively, and the correlations were assessed using Spearman’s rank correlation coefficients.

Results: When tested in 134 HBV patients, the test-retest reliability was supported with all scales showing acceptable correlation coefficients (i.e., α > 0.7). Item-to-scale correlations were good with most items highly correlated with their hypothesized scales. Convergent and divergent construct validities were supported by the hypothesized correlations between the HQLQ and the EQ-5D domains.

Conclusions: The culturally adapted questionnaire has good validity and reliability for use in Singapore.

Keywords: disease-specific instrument, EQ-5D, health-related quality of life, outcomes research, SF-36.

Introduction

Currently, Health-related Quality of Life (HRQoL) instruments are not routinely used in clinical settings because of various reasons; one of them is the lack of validation study. Recently, we published two studies of HRQol in patients with chronic hepatitis B using both generic and disease-specific instruments; results from both studies showed a gradual deterioration in HRQoL with progression of liver disease [1,2].

We adapted the Hepatitis Quality of Life Questionnaire (HQLQ) [2], which covers hepatitis specifically and contains the SF-36 health survey as the generic core [3]. The major adaptation of the English version was the changes of English terms not commonly used in Singapore. For example, “weighted down” in the original questionnaire are seldom used locally, and has been replaced with “down” or depressed which are more colloquial and well accepted in Singapore. Similarly, “a wonderful adventure” in the original HQLQ was replaced with “exciting” as the original phrase is found to be abstract by focus group participants.

As Singapore is a multiracial country with a sizeable ethnic Chinese population (75%), validating a Chinese version of the HQLQ for those who cannot understand English well would be necessary. Therefore, the main objective of this study was to validate a Chinese version of the HQLQ for use in Chinese-speaking hepatitis B patients in Singapore.

Methods

Study Design

A cross-sectional, multi-phase, institutional review board approved study was conducted at the National University Hospital (NUH) with informed consent obtained from all participating subjects.

Translation and Cultural Adaptation of the Instrument

As the Singapore Chinese version of the SF-36 has been validated, Chinese translation of the original English version of the HQLQ was only performed for the additional scales from the HQLQ (five generic items, two generic scales: health distress and positive well-being, as well as hepatitis-specific limitations and hepatitis-specific health distress scales) according to the standard guidelines suggested by Guillemin [4] as follows: forward translations were performed independently by two bilingual translators; reconciliation of both forward versions; another two translators independently translated the reconciled Chinese version back into English. The final version was obtained after a pilot testing of the revised questionnaire in eight local Chinese-speaking individuals.

Validation Study

Patient recruitment. The finalized Chinese version of HQLQ was validated in a sample of hepatitis B virus (HBV) patients attending outpatient clinics at NUH. Other inclusion criteria were above age 18 and ability to self-complete the questionnaires in Chinese. Aside from providing their socio-demographic information, participants were asked to fill in the HQLQ and EQ-5D questionnaires before or after their clinic appointment.

For test-retest reliability, HBV patients were given a second set of the questionnaire, and were asked to complete the
questionnaire at least 3 days from their first test and return the questionnaire using a return-business envelope.

### Psychometric properties analyses
Reliability was assessed by Cronbach’s alpha coefficients (for internal consistency) and intra-class correlation coefficients (for test-retest reliability) [5,6]. For interpretation, Cronbach’s alpha coefficients above 0.7 and 0.9 are generally regarded as acceptable for conducting group comparisons and assessing individual subjects, respectively [6].

Scoring assumptions of the HQLQ were investigated at item level by examining item-scale convergent validity [7]. Item-to-scale correlation (corrected for overlap) was assessed using Spearman’s rank correlations ($r$). Item-to-scale correlation was considered as strong correlation, $0.35$ to $0.5$ as moderate correlation, and $0.2$ to $0.34$ as weak correlation [10]. The scale correlation (corrected for overlap) was assessed using Spearman’s rank correlations ($r$) between scale scores and their constituent items with $p \geq 0.4$ considered as acceptable [5]. The scale score distributions were evaluated by computing the percentage of respondents achieving either the highest possible score (ceiling) or the lowest possible score (floor) [3].

Construct validity of the HQLQ at scale level was investigated by examining the correlations between HQLQ and EQ-5D [8] domains. As the SF-36 has been validated previously in Singapore [9], only those additional HQLQ scales were tested. Convergent and divergent validities were assessed using Spearman’s rank correlation coefficients ($p$), with a rho value >0.5 considered as strong correlation, 0.35 to 0.5 as moderate correlation, and 0.2 to 0.34 as weak correlation [10].

Two other a priori hypotheses based on clinical expectations were generated for divergent construct validity where moderate-to-strong correlations (i.e., correlation coefficient $\geq0.35$) were expected between domains measuring similar constructs, namely: 1) HQLQ Health Distress, Positive Well-being, and Hepatitis-specific Health Distress with EQ-5D Anxiety/depression; and 2) HQLQ Hepatitis-specific Limitations with EQ-5D Mobility and Usual Activities. Another two a priori hypotheses were generated for divergent construct validity where weak correlations were expected between domains measuring dissimilar constructs, namely: 1) HQLQ Health Distress, Positive Well-being, and Hepatitis-specific Health Distress with EQ-5D Mobility, Self-care, Usual Activities, and Pain/discomfort; and 2) HQLQ Hepatitis-specific Limitations with EQ-5D Anxiety/depression and Self-care.

All statistical analyses were performed using SPSS for Windows version 13.0 (SPSS Inc., Chicago, IL). Statistical significance for all tests was set at 0.05.

### Results

#### Cultural Adaptation of Instrument
The Chinese version of the culturally adapted HQLQ was well accepted in the pilot testing; the respondents generally found the questionnaire clear and easy to understand and to answer. Hence, the adapted questionnaire was used in the subsequent validation study without any further revisions.

#### Sample Characteristics for Cross-Sectional Validation Study
From November 2003 to November 2006, all Chinese outpatients with chronic hepatitis B infections were approached to participate in this study. In total, 134 chronic hepatitis B patients (47 asymptomatic carrier, 33 chronic hepatitis B, 24 compensated cirrhosis, 9 decompensated cirrhosis, 9 hepatocellular carcinoma, and 12 post-liver transplants) completed the questionnaires. The percentage of nonresponders and patients with reported comorbidities was less than 1% and 6%, respectively, so the data have not been analyzed further. The average time taken to complete the questionnaire was 20 minutes (range: 10–30 minutes). The mean age of the participants was 50.7 years (SD: 12.3; range of 22–80 years) with 64.2% being male. Additional socio-demographic information will be made available upon request.

#### Psychometric Properties
The descriptive statistics and reliability coefficient (Cronbach’s alpha) of the scales were detailed as per Table 1. Missing data for each scale were low (<2%). Ceiling effect was observed for most scales with Role Physical and Role Emotional scales showing the highest percentage (63.4%) of ceiling effect, respectively. The other notable ceiling effects were observed in the Hepatitis-specific Limitation, Bodily Pain, and Physical Functioning scales (44.8%, 44.0%, and 40.3%, respectively). Although floor effects were observed in six scales, the percentage of respondents scoring the floor in these scales was generally less than 4%. The internal consistency reliability coefficients, alpha values were good with $\alpha > 0.7$ in all the scales and half of these were $> 0.8$. Notably, alpha values for all the HQLQ-specific component scales were even $> 0.9$.

A total of 28 subjects participated in the test-retest analysis. Acceptable correlation coefficients (i.e., $\alpha > 0.7$) were observed in all the scales, with half of these showed a value greater than 0.8 (Table 2).

Item-to-scale correlations were good with all items highly correlated with their hypothesized scales, with only one item each from Physical Functioning and Role Emotional scales ($p = 0.21$ and 0.36, respectively) below the cutoff value ($p \geq 0.4$) recommended for scale construction (Table 2).

For convergent construct validity, moderate-to-high correlations were presented for six of nine subhypotheses ($p$ between

### Table 1: Descriptive statistics and reliability for each scale of Hepatitis Quality of Life Questionnaire (HQLQ; n = 134)

<table>
<thead>
<tr>
<th></th>
<th>Number of items</th>
<th>Missing data (%)</th>
<th>Minimum score</th>
<th>Maximum score</th>
<th>% showing floor effect</th>
<th>% showing ceiling effect</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36 components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>10</td>
<td>2.2</td>
<td>28.57</td>
<td>100.0</td>
<td>0.0</td>
<td>40.3</td>
<td>0.864</td>
</tr>
<tr>
<td>Role physical</td>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>3.7</td>
<td>63.4</td>
<td>0.830</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>4</td>
<td>0.7</td>
<td>22.0</td>
<td>100.0</td>
<td>0.0</td>
<td>44.0</td>
<td>0.720</td>
</tr>
<tr>
<td>General health</td>
<td>5</td>
<td>1.5</td>
<td>0.0</td>
<td>100.0</td>
<td>0.7</td>
<td>0.7</td>
<td>0.750</td>
</tr>
<tr>
<td>Vitality</td>
<td>4</td>
<td>0.7</td>
<td>5.0</td>
<td>95.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.764</td>
</tr>
<tr>
<td>Social functioning</td>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>87.5</td>
<td>0.7</td>
<td>0.0</td>
<td>0.775</td>
</tr>
<tr>
<td>Role emotional</td>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>3.0</td>
<td>63.4</td>
<td>0.777</td>
</tr>
<tr>
<td>Mental health</td>
<td>5</td>
<td>0.7</td>
<td>8.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.798</td>
</tr>
<tr>
<td>HQLQ-specific component scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health distress</td>
<td>4</td>
<td>0.7</td>
<td>0.0</td>
<td>100.0</td>
<td>1.5</td>
<td>20.9</td>
<td>0.964</td>
</tr>
<tr>
<td>Positive well-being</td>
<td>4</td>
<td>0.7</td>
<td>5.0</td>
<td>100.0</td>
<td>0.0</td>
<td>3.7</td>
<td>0.929</td>
</tr>
<tr>
<td>Hepatitis-specific limitations</td>
<td>3</td>
<td>0.7</td>
<td>26.7</td>
<td>100.0</td>
<td>0.0</td>
<td>44.8</td>
<td>0.912</td>
</tr>
<tr>
<td>Hepatitis-specific health distress</td>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>1.5</td>
<td>31.3</td>
<td>0.966</td>
</tr>
</tbody>
</table>
The construct validities of the instrument are supported by convergent and divergent construct validity as measured by the correlation coefficient of the two a priori hypotheses. 0.39 and 0.51 of the two a priori hypotheses. On the other hand, divergent construct validity was supported by all a priori hypotheses, with all scales from HQLQ correlated weakly (p between 0.01 and 0.27) with EQ-5D domains measuring dissimilar constructs (data will be made available upon request).

Discussions

To the best of our knowledge, this is the first Chinese disease-specific HRQoL instrument adapted and validated for HBV patients in Asia.

The significant ceiling effect observed in the present study especially in Role Physical and Role Emotional scales were also reported previously by Bayliss et al. [3] and our validation study of the English version of this questionnaire [2]. Nevertheless, this effect might be overcome by substituting the current dichotomous response choices with five level response options used in SF-36 V2 or HQLQ™ (Version 2) [11,12].

In the current study, the internal consistency reliability coefficients, Cronbach’s alpha, of the multi-item scales were good with α > 0.7 in all the scales. This is consistent with the results achieved by our previous validation study of the English version and supports the assertion that the Chinese version of HQLQ has good internal consistency reliability. Test-retest reliability was also strongly supported with all the 12 scales showing acceptable correlation coefficients (i.e., α > 0.7).

In the measurement of item-to-scale correlations, one item each from the Physical Functioning and Role Emotional scales showed correlation coefficient below the cutoff value. The item from Role Emotional scale also reported the below cutoff value in our previous validation study for the English version of HQLQ [2]. Likewise, this item was also shown to have a relatively low correlation with its parent scale in the previous validation study in hepatitis C patients, probably due to its low rates of endorsement [3]. Nevertheless, as correlation coefficient of 0.36 from this item is close to the recommended cutoff value, i.e., ρ > 0.4, more validation studies are needed to allow a conclusive decision whether this item should be removed or moved to other scales.

The construct validities of the instrument are supported by the expected correlations between the HQLQ scales and the EQ-5D domains measuring similar or dissimilar constructs, again consistent with our validation study of the adapted English version of HQLQ [2].

Finally, we acknowledged several limitations of the present study, the first being the cross-sectional study design of the study which precluded the evaluation of the instrument’s responsiveness. In addition, background information of nonrespondents was not collected in this study. Nevertheless, given the low non-response rate of this study (<1%) and the consistency in response from the groups, it seemed that our sample would be a reasonably good representation of the target population.

Conclusions

Our study found evidence to support the current Chinese version of HQLQ as culturally appropriate, valid, and reliable in a sample of Chinese-speaking Singapore residents with various stages of HBV infection.

Acknowledgments

We would like to express our gratitude to research nurses Amelia Cheok, Belinda Mak and Winnie Chua for their assistance during data collection. We would also like to thank all of the nurses and staffs from Unit Digestive Centre, National University Hospital for their help and cooperation in facilitating the study.

Source of financial support: This study was funded by NHG grant NHG-RPR/01086 and National University of Singapore Faculty Research Fund.

References