Health-related quality of life questionnaires used in primary biliary cholangitis: a systematic review

Xin Ai¹, Xian Yang¹, Jia-min Xu¹, Wen-xia Yang¹, and Ying-Mei Tang²

¹The Second Affiliated Hospital of Kunming Medical University
²The Second Affiliated Hospital of Kunming Medical University, Yunnan Research Center for Liver Diseases

November 30, 2020

Abstract

Objective The purpose of this systematic review was to assess the suitability of HRQOL questionnaires in patients with primary biliary cholangitis. Methods Five electronic databases were searched. The validity of translated questionnaires, floor and ceiling effects, internal consistency and test-retest reliability were investigated. Results Forty-four studies were included, of which fifteen HRQOL questionnaires were identified. The instruments used most frequently were the PBC-40 (n = 22), followed by the SF-36 (n = 19), PBC-27 (n = 4), CLDQ (n = 3) and NIDDK-QA (n = 2), the remaining instruments were used only once. Twenty-six studies used a translated HRQOL questionnaire and only six reported or referenced a validation of the translated questionnaire. Conclusions PBC-specific HRQOL questionnaires used in primary biliary cholangitis have generally good psychometric properties. But lots of studies directly applied the HRQOL tools without verifying the HRQOL tools validity and reliability in PBC patients. Thus, it is better for clinicians and researchers to test the measurement properties of HRQOL questionnaires before use it.
Key words: review, primary biliary cholangitis, health-related quality of life

*Corresponding author: Ying-mei Tang
mail address: tangyingmei_med@kmmu.edu.cn

Review criteria:
Inclusion criteria
1. Original studies of adult patients (≥18 years old) with primary biliary cholangitis, studies reporting generic or disease-specific HRQOL questionnaires as outcome measure.
2. The study had to be a randomized controlled trial, cohort study, case control study, clinical trial, or validation study of HRQL instruments.

Message for the clinic:
PBC-specific HRQOL questionnaires used in primary biliary cholangitis have generally good psychometric properties. It is better for clinicians and researchers to test the measurement properties of HRQOL questionnaires before use it.

1.Introduction
Primary biliary cholangitis (PBC) is an autoimmune disease, which characterized by interlobular cholangitis with progressive destruction, cholestasis, fibrosis and may develop into end-stage liver disease eventually\(^1\), \(^2\). High-titer specific anti-mitochondrial antibodies and anti-nuclear antibodies GP120 and SP100 appeared in serum of PBC patients, and the level of IgM in peripheral blood also increased correspondingly\(^3\). Fatigue and pruritus are the most common symptoms of primary biliary cirrhosis, which affects patients’ quality of life seriously. Ursodeoxycholic acid (UDCA) is the first valid drug in the treatment of PBC, but 30% of patients with PBC do not have a biochemical response to UDCA\(^4\), \(^5\). The World Health Organization (WHO) defines quality of life as individuals’ perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns\(^6\). And health-related quality of life (HRQOL) is defined as the influence of health and disease on quality of life\(^7\). Health-related quality of life (HRQOL) measures in chronic disease management is necessary as it assesses the overall impact on health from the patient’s perspective. A range of different questionnaires are used to evaluate HRQOL in PBC, including disease-specific and generic questionnaires, but no guidance of HRQOL questionnaires properties for clinicians and researchers to choose. The aim of this review was to evaluate the suitability of questionnaires used to assess HRQOL in primary biliary cholangitis.

2.Materials and methods
We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines for systematic reviews\(^8\) to ensure a complete reporting in the development of the study protocol. This systematic review was performed in January 2020.

2.1 Data sources
We performed electronic searches in the following five databases: PubMed, the Cochrane Library, Web of Science, Wan Fang Database, and Chinese National Knowledge Infrastructure. The search was run on 17 January 2020.

2.2 Search strategy
The search strategy included a combination of the following relevant keywords: “health-related quality of life”, “quality of life”, “primary biliary cholangitis” and “primary biliary cirrhosis”. Electronic database searches were limited to English-language publications.

2.3 Inclusion and exclusion criteria
Inclusion criteria

1. Original studies of adult patients (≥18 years old) with primary biliary cholangitis, studies reporting generic or disease-specific HRQOL questionnaires as outcome measure.

2. The study had to be a randomized controlled trial, cohort study, case control study, clinical trial, or validation study of HRQL instruments.

3. Articles published between January 1999 and January 2020 (The publication period of 20 years considered for this systematic review is defined based on the reference list from Osborne et al.)[9].

Exclusion criteria

1. Studies reporting quality of life for primary sclerosing cholangitis, autoimmune hepatitis.

2. Case reports, conference abstracts, case series, and dissertations were excluded.

3. Using fatigue scales or pruritus scales to assess patients’ quality of life.

4. Number of included cases ≤10.

2.4 Data extraction

Two reviewers (X.Y. and J.-M.X.) assessed the full-text publications and extracted data using standardized electronic forms independently. Disagreement between the reviewers about eligibility was resolved through discussion. Data extracted from the selected studies including: author, year of publication, study design, aim of the study, study characteristics, intervention, HRQOL tools, results. For translated questionnaires, we evaluated floor and ceiling effects, missing data, internal consistency and test-retest reliability. Finally, we performed a quality analysis of the included studies following the Hawker’s tool, which provides clear description of ratings, that is, “good,” “fair,” “poor” and “very poor,” and has been designed to assess quality of studies covering a variety of research paradigms[10].

3. Results

3.1 Study selection

The search retrieved 1175 publications and 435 full-text records were obtained after duplicates removed. Studies were published between 1999.1 and 2020.1. 391 studies were excluded with reasons. Finally, 44 studies met the inclusion criteria. A PRISMA flow chart illustrates the studies selection process and reasons for exclusion (Figure 1).

3.2 Overview of included studies

A total of 15 HRQOL questionnaires were used in the included 44 studies, descriptions of the used questionnaires are presented in table 1. The instruments used most frequently were the PBC-40 (n = 21), followed by the SF-36 (n = 19), PBC-27 (n = 4), CLDQ (n = 3) and NIDDK-QA(n = 2), the remaining instruments were used only once (Supplementary table 1-3). These instruments were initially designed for various purposes and referred to different concepts. The PBC-40 was the first questionnaire for use in patients with PBC specifically. The SF-36 is a standardized measure of health status, commonly used as a HRQOL questionnaire. The PBC-27 and the PBC-10 were brief vision of PBC-40. The CLDQ, CLDQ-I and LDSI2.0 were liver disease-specific questionnaires. The other questionnaires were developed to measure generic quality of life.

Validity of translated questionnaires

Twenty-six studies used a translated HRQOL questionnaire and only six reported or referenced a validation of the translated questionnaire[11-16].

Floor/ceiling effects and missing data
Floor and ceiling effects were reported for only two HRQOL questionnaires (PBC-40 and PBC-10). There were no significant ceiling effects but a moderate floor effect was found for the itch domain in PBC-40. The PBC-10 did not have a significant ceiling effect but flooring effects were observed in all 10 items similarly\(^{[15, 17]}\).

Internal consistency

Cronbach’s \(\alpha\) coefficients for the HRQOL questionnaires ranged from adequate to excellent. The PBC-40 had the highest internal consistency (Cronbach’s \(\alpha\) coefficient 0.72–0.95)\(^{[17]}\). PBC-27 Cronbach’s \(\alpha\) coefficients ranged from 0.45 to 0.93\(^{[13]}\), NIDDK-QA from 0.87 to 0.94 and for PBC-10 it was 0.90\(^{[15, 16]}\). The internal consistency of other questionnaires was not reported.

Test-retest reliability

Test-retest reliability data were available for PBC-40, NIDDK-QA and PBC-10. The test-retest reliability of NIDDK-QA was measured by Pearson Correlation, it was ranged from 0.82 to 0.99. Test-retest reliability was calculated using intraclass correlation coefficients (ICCs) of PBC-40 and PBC-10. The PBC-40 ICCs ranged from 0.83 to 0.96 and for PBC-10 it was 0.945.

---

**Figure 1** PRISMA flow diagram of the study selection process.

**Table 1** Generic quality of life questionnaires and disease-specific quality of life questionnaires

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Scale (item)</th>
<th>Domains description</th>
<th>Scoring/administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease-specific quality of life questionnaires</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaires</td>
<td>Scale (item)</td>
<td>Domains description</td>
<td>Scoring/administration</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>PBC-40[^17]</td>
<td>6(40)</td>
<td>Fatigue, emotional, social, cognitive, itch and symptoms</td>
<td>5 point scale, 1(not at all) to 5(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>PBC-27[^12]</td>
<td>7(27)</td>
<td>Other Symptoms, dryness, itch, fatigue, cognitive, emotional and social</td>
<td>5 point scale, 1(not at all) to 5(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>PBC-10[^15]</td>
<td>6(10)</td>
<td>Fatigue, emotional, social, cognitive, itch and symptoms</td>
<td>5 point scale, 1(not at all) to 5(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>CLDQ[^18]</td>
<td>6(29)</td>
<td>Abdominal symptoms, fatigue, systemic symptoms, activity, emotional function, worry</td>
<td>7 point scale, 1(not at all) to 7(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>CLDQ-I[^19]</td>
<td>6(28)</td>
<td>Abdominal symptoms, fatigue, systemic symptoms, activity, emotional function, worry</td>
<td>6 point scale, 1(not at all) to 7(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>LDSI 2.0[^20]</td>
<td>18</td>
<td>Itch, joint pain, pain in the right upper abdomen, sleepiness during the day, worry about family situation, decreased appetite, depression, fear of complications and jaundice, nine other items measure the hindrance of these symptoms to daily activities</td>
<td>5 point scale, 1(not at all) to 5(to a high extent) Subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>Generic quality of life questionnaires</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36[^21]</td>
<td>8(36)</td>
<td>Physical functioning, physical role functioning, limitation emotional role limitation, bodily pain, mental health, social functioning, vitality and general health perception</td>
<td>100 point scale, 0(not at all) to 100(very much) Total and subscale score Higher score indicates greater health</td>
</tr>
<tr>
<td>SF-12[^22]</td>
<td>2(12)</td>
<td>Physical component and mental summary</td>
<td>100 point scale, 0(not at all) to 100(very much) Total and subscale score Higher score indicates greater health</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>Scale (item)</td>
<td>Domains description</td>
<td>Scoring/administration</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>15D(generic)[23]</td>
<td>15</td>
<td>Breathing, mental function, speech (communication), vision, mobility, usual activities, vitality, hearing, eating, elimination, sleeping, distress, discomfort and symptoms, sexual activity, and depression</td>
<td>5 point scale, 1(not at all) to 5(very much) Total score Higher score indicates poorer health</td>
</tr>
<tr>
<td>NIDDK LTD-QOL[24]</td>
<td>5(21)</td>
<td>Measures of disease, psychological status, personal function, social and role function, and general health perception</td>
<td>5 point scale, 0(not at all) to 4(extremely) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>NIDDK-QA[16]</td>
<td>4(47)</td>
<td>Liver disease symptoms, physical functioning, health satisfaction and overall well-being</td>
<td>5 point scale, 0(not at all) to 4(extreme) Total and subscale score Higher score indicates greater health</td>
</tr>
<tr>
<td>HUI-Mark-2[25]</td>
<td>7(15)</td>
<td>Sensation, mobility, emotion, cognition, selfcare, pain and fertility</td>
<td>2 point scale, 0(not at all) to 1(very much) Total and subscale score Higher score indicates greater health</td>
</tr>
<tr>
<td>HUI-Mark-3[26]</td>
<td>8(15)</td>
<td>Vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain</td>
<td>2 point scale, 0(not at all) to 1(very much) Total and subscale score Higher score indicates greater health</td>
</tr>
<tr>
<td>NHP[27]</td>
<td>6(38)</td>
<td>Energy, sleep, pain, emotional reactions, social isolation, and physical mobility</td>
<td>2 point scale, 0(no) to 1(yes) Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>NHANES III</td>
<td>4(83)</td>
<td>Articular symptoms, activity limitations (in both household and job-related activities), hobbies and leisure activities, and social life</td>
<td>Total and subscale score Higher score indicates poorer health</td>
</tr>
<tr>
<td>PROMIS-29(generic)[28]</td>
<td>7(29)</td>
<td>Physical function, depression, anxiety, fatigue, sleep disturbance, ability to participate in social roles and activities (social roles), pain interference, and one pain intensity item</td>
<td>5 point scale, 1(not at all) to 5(very much) Total and subscale score Higher score indicates poorer health</td>
</tr>
</tbody>
</table>
Table 2 Summary of psychometric properties of the identified HRQOL questionnaires

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Validity</th>
<th>Reliability</th>
<th>Responsiveness and floor/ceiling effects</th>
<th>Minimal important difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PBC-40</strong></td>
<td>Construct validity Moderate to high correlations with the SF-36 and ESI-55 scales, except for “Physical functioning” domain of the SF-36 and “Itch” domain of the PBC-40</td>
<td>Internal consistency Cronbach’s alpha &gt;0.7 for all scales. Test-retest ICC ranged 0.83-0.96</td>
<td>Floor and ceiling effects There were no ceiling effects but there was a noticeable floor effect in the itch domain (36.7%)[15]</td>
<td>The minimal important difference was 3.539[15]</td>
</tr>
<tr>
<td><strong>PBC-27</strong></td>
<td>Convergent validity minor correlations between the PBC-27 social factor and the SF-36 social functioning, but moderate to high correlations in majority of scales</td>
<td>Internal consistency Cronbach’s alpha &gt;0.7 for all scales, except for domains “Dryness”, “Symptoms” and “Fatigue”.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PBC-10</strong></td>
<td>Content validity Generally good, developed by hepatology experts in PBC</td>
<td>Internal consistency was excellent with Cronbach’s alpha of 0.936. Test-retest ICC=0.945 for all scales</td>
<td>Floor and ceiling effects There were no ceiling effects but a floor effect was obvious. Responsiveness ratio of 0.38, effect size of 0.39 and standardised response mean of 1.25.</td>
<td></td>
</tr>
<tr>
<td><strong>NIDDK-QA</strong></td>
<td>Concurrent validity Small to high correlations with the SF-36 scales, P &lt;0.01</td>
<td>Internal consistency Cronbach’s alpha ranged from 0.87 to 0.94 Test-retest Pearson coefficients: 0.82-0.99, P&lt;0.01</td>
<td>Responsiveness Small to moderate responsiveness</td>
<td></td>
</tr>
</tbody>
</table>
4. Discussion

This is the first systematic review of the suitability of HRQOL questionnaires used to assess primary biliary cholangitis, with the goal to provide evidence for clinicians and researchers concerning its use within clinical practice and research. We explored the suitability of these instruments through their psychometric properties and their content. There are 15 HRQOL questionnaires currently available.

First, this review found that a high degree of heterogeneity in HRQOL used instruments in the primary biliary cholangitis studies: fifteen different types of instruments were used in forty-four included articles. Among these forty-four included articles, only three articles were randomized and among non-randomized controlled trials, only eight assessed HRQOL as a primary endpoint. This finding shows that the relevance of HRQOL is not fully recognized in primary biliary cholangitis studies. If we consider observational studies, few studies were longitudinal; the others were cross-sectional with no possibility of exploring a change over time.

Many studies used translated HRQOL questionnaires. While there were good examples of translated validity in PBC-40, there were many studies where this procedure was either not conducted or reported. It is essential that HRQOL questionnaires are translated and validated using well recognized and standardized procedures to ensure that they are appropriately adapted to accommodate cultural differences. Since one’s functional status, such as language barriers, quality of education, literacy level will have a direct effect on HRQOL scores. Strengthening investigators’ ability to interpret results and formulating clinically relevant conclusions when using HRQOL questionnaire will ameliorate variability of scores. Inductive method was performed to assess psychometric properties of instruments in this review, minimal important difference, responsiveness and floor/ceiling effects, test-rest reliability, internal consistency (Cronbach’s alpha) and various forms of construct validity (convergent, content and concurrent ) were assessed. Except PBC-10, no other questionnaire provides complete psychometric validation in patients with PBC. Another common problem of most studies was that no sample size calculation was provided, which offers stronger implications on validity.

Our review highlights the questionnaire options available, reminding researchers should take into account psychometric properties and content when choose which HRQOL is the most appropriate to use in primary biliary cholangitis. The PBC-40 was the most widely studied questionnaire, with most extensive psychometric evaluation, which is the first disease-specific HRQOL assessment tool for PBC patients. Cronbach’s $\alpha$ scores in all six domains were above 0.7, three of the domains (fatigue, cognitive, and itch) met the recommended minimum of 0.7 for individual patient comparisons. Test-retest reliability scores were above 0.8. PBC-40 is a good choice for a research study if extensive experience of use of an HRQOL questionnaire is important. Its disadvantage is that it is too lengthy and time consuming to use in the normal clinical practice. The next most commonly used tool was the SF-36. Of generic quality of life questionnaires, the SF-36 performed best in terms of internal consistency and test-retest reliability. Though SF-36 provides a generic HRQOL assessment tool, it is less sensitive to impairments in HRQOL caused by pruritus and did not explore the quality of sleep and quality of wake domains. Briefer tools with simplified scoring for use in the clinical and research settings are needed urgently. The PBC-27 is one such example, which was shown to be equally effective in detecting the impact of PBC on HRQOL. However, the PBC-27 has potential language problems, since its original version was developed for use in Italian and Japanese patients with primary biliary cirrhosis. Evaluations of PBC-27 in Poland were performed after translating it into Polish. Thus, the PBC-27 requires testing of its measurement stability for the different language versions. Another short HRQL questionnaire, PBC-10, had good internal consistency (Cronbach’s $\alpha$ 0.905) and test-retest reliability. PBC-10 demonstrated no ceiling effects but a floor effect was noted, which requires further study. The CLDQ, a liver disease-specific HRQOL questionnaire, was developed to measure longitudinal change in individuals with chronic liver disease, although it didn’t cover pruritus and fatigue domains in PBC patients well. The CLDQ-I (CLDQ Italy vision) and LDSI 2.0 are liver disease-specific HRQOL questionnaires. However, the LDSI 2.0 lacks comprehensiveness as a QOL instrument in terms of emotional well-being/emotional functioning and, if used, should be used in conjunction with other measures of well-
being. The SF-12, 15D, NIDDK LTD-QOL, NIDDK-QA, HUI-Mark-2, HUI-Mark-3, NHP, NHANES III and PROMIS-29 are generic HRQOL tools. The psychometric properties of the above tools has not been verified in PBC patients(except for NIDDK-QA), thus it is not recommended use them in PBC patients.

There are some limitations in our systematic review. First, PRISMA is just one methodology that can be used to synthesis or evaluate outcome measures and other methods might be equally valid or provide different perspectives. Secondly, we did not search the gray literature (inability to identify ongoing or unpublished studies), which produces selection bias. And our review was limited to English-language articles, some studies may not have been included due to this.

5. Conclusion

PBC-specific HRQOL questionnaires used in primary biliary cholangitis have generally good psychometric properties. But lots of studies directly applied the HRQOL tools without verifying the HRQOL tools validity and reliability in PBC patients. Investigators should select questionnaires for their study based on the ease of administration and the questionnaire’s correlation with the primary health domain under investigation. Thus, more robust psychometric design studies to test the measurement properties of HRQOL questionnaires as the primary outcome of investigation are highly needed.

Conflict of interest

The authors have no conflict of interest to declare.

Contributions

X.A. contributed significantly to the conception and design of the study, data extraction, critical appraisal, interpretation of data and drafting of the manuscript. X.Y. and J.-M.X. were involved in literature search, critical appraisal and interpretation of data and drafting. W.-X.Y. was involved in critical appraisal and drafting. Y.-M.T. was involved in the drafting and review of the manuscript. All authors have given their final approval on the manuscript to be published.

Funding

This work was supported by grants from the National Natural Science Foundation of China (NO.81660102), Project of Science and Technology Innovation team in Colleges and Universities in Yunnan Province, Yunnan Natural Science Foundation (NO.201701UH00618), Yunnan Health Science and Technology Program Project (NO.2017NS280) and Yunnan Provincial Science and Technology Department - Kunming Medical University applied basic research joint special fund project (2018FE001(-051)).

Patient consent for publication

Not required.

Copyright permissions

All figures and tables were original.

References


Hosted file
