BMJ Open Gastroenterology

Measurement properties of the PBC-40 and PBC-27: a Dutch validation study

Rozanne C de Veer ,[1] Geraldine da Silva,[1] Maria C van Hooff,[1] Maren H Harms,[1] Herold J Metselaar,[1] José Willemse,[2] Elaine Utomo,[3] Adriaan J van der Meer[1]

To cite: de Veer RC, da Silva G, van Hooff MC, *et al*. Measurement properties of the PBC-40 and PBC-27: a Dutch validation study. *BMJ Open Gastro* 2021;**8**:e000758. doi:10.1136/bmjgast-2021-000758

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgast-2021-000758).

Received 4 August 2021
Accepted 28 November 2021

[1]Gastroenterology and Hepatology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands
[2]Dutch Liver Patients Association, Hoogland, The Netherlands
[3]Independent researcher, Berkel en Rodenrijs, The Netherlands

Correspondence to
Adriaan J van der Meer;
a.vandermeer@erasmusmc.nl

ABSTRACT

Objective Patients with primary biliary cholangitis (PBC) have an impaired health-related quality of life (HRQoL). Practice guidelines recommend evaluating the HRQoL in all patients with PBC. The aim of this study was to assess the reliability and validity of our Dutch translation of the PBC-40, a PBC-specific measure of the HRQoL.

Design The PBC-40 was translated into Dutch following standardised forward–backward procedures. Participants received the Dutch PBC-40 and the RAND-36 (a validated Dutch version of the 36-Item Short Form Health Survey) through postal mail. The PBC-27 is an abridged version of the PBC-40. Internal consistency between the items within the PBC-40/PBC-27 domains was assessed by Cronbach's alpha. In addition, score distributions were analysed on floor and ceiling effects. Construct validity was assessed by hypotheses testing using Pearson's correlation between the PBC-40/PBC-27 domains and RAND-36 scales.

Results 177 patients with PBC were included. The mean age was 61.1 (SD 9.9) years and the majority of patients was female (n=164, 92.7%). From the 7080 PBC-40 items, 61 items (0.9%) were missing and 342 items (4.8%) were answered with the 'does not apply' option. Each PBC-40 domain had a Cronbach's α of >0.70, with the highest in the domain fatigue (0.95). For the PBC-27, the lowest Cronbach's α was 0.69. Floor effects were present in three domains (cognition 19.3%, itch 27.0% and social 25.0% (only for PBC-27)). No ceiling effects were observed. All domains were significantly correlated with the corresponding RAND-36 scale(s) ($p<0.001$ for all). The strongest correlation was between the PBC-40 domain fatigue and the RAND-36 vitality scale ($r=-0.834$).

Conclusion Our findings demonstrate the reliability and validity of the Dutch PBC-40 and PBC-27 for the assessment of the HRQoL in patients with PBC. This PBC-specific measure can be used in Dutch-speaking patients with PBC for both research and clinical purposes.

Summary box

What is already known about this subject?

► The health-related quality of life (HRQoL) of patients with primary biliary cholangitis (PBC) is impaired and requires assessment.

What are the new findings?

► The reliability and validity of the Dutch translation of the PBC-40, the only available PBC-specific quality of life measure, are adequate.

How might it impact on clinical practice in the foreseeable future?

► The Dutch PBC-40 is a useful PBC-specific HRQoL measure and can be used for both research and clinical purposes.

INTRODUCTION

Primary biliary cholangitis (PBC) is a chronic cholestatic liver disease with autoimmune features, in which the small intrahepatic bile ducts are slowly but progressively destroyed.[1] The disease may progress towards cirrhosis, at which stage patients are at risk of liver failure and hepatocellular carcinoma. The health-related quality of life (HRQoL) of patients with PBC has proven to be significantly impaired due to hepatic dysfunction and other extrahepatic disease manifestations.[2 3] Complaints of patients with PBC extend beyond classical extrahepatic symptoms such as fatigue, pruritus or sicca syndrome, and may include cognitive dysfunction, bone/joint pain or restless legs. The severity of these symptoms is independent of the histological or biochemical stage of the disease.[4]

The European Association for the Study of the Liver (EASL) recommends evaluation of all patients with PBC on the presence of symptoms followed by a quantification of their impact on the HRQoL.[5] However, measuring HRQoL tends to be complex due to its subjective and multifaceted nature. In 2005, the UK-PBC group developed the PBC-40, the first and only PBC-specific quality of life measure.[6] In order to use a translation of the PBC-40, it is essential to assess its validity and reliability as several studies have shown that a measure that is valid and reliable for a particular language and culture may not prove so when used in a different population.[7] Over recent years, the PBC-40 has been cross-culturally adapted and translated into different languages.[8–10] Subsequently, a shorter version of the PBC-40, the PBC-27, was created and validated.[8] In the

Table 1 Cohort characteristics

	N=177
Age (years)*	61.1 (9.9)
Female, n (%)	164 (92.7)
Cirrhosis, n (%)†	28/175 (16.0)
Hospitals	
University centres	93 (52.5)
Non-university centres	84 (47.5)
Duration of disease	
Less than 12 months	5 (2.8)
Between 1 and 5 years	37 (20.9)
Between 5 and 10 years	38 (21.5)
Longer than 10 years	97 (54.8)
Educational level	
Primary education	3 (1.7)
Secondary education	28 (15.9)
Intermediate vocational education	72 (40.9)
High vocational education	46 (26.2)
University	21 (11.9)
Other	6 (3.4)
UDCA treatment, n (%)	166 (93.8)
Pruritus VAS score‡	1.0 (0.0–4.0)

*Data are expressed as mean with SD.
†Data on cirrhosis are missing for two patients.
‡Data are expressed as median with IQR.
UDCA, ursodeoxycholic acid; VAS, Visual Analogue Scale.

Netherlands, there is currently no PBC-specific quality of life measure available. A standardised assessment of the HRQoL of Dutch-speaking patients with PBC would facilitate routine screening in clinical practice and contribute to ongoing and future research, both nationally and internationally. Therefore, the aim of this study was to assess the reliability and validity of our Dutch translation of the PBC-40 and the PBC-27 in a population of patients with PBC in the Netherlands.

PATIENTS AND METHODS

Study population and design

All members of the Dutch Liver Patients Association (DLPA) who are registered with a diagnosis of PBC and patients with PBC of the outpatient clinic of the department of gastroenterology and hepatology of the Erasmus University Medical Centre Rotterdam diagnosed with PBC received an invitation between August and October 2020 to participate in this study. Patients were eligible for inclusion in case of an established PBC diagnosis according to the EASL guideline.[5] Patients received two self-report questionnaires, the Dutch PBC-40 and the RAND-36, through postal mail. Demographic and clinical questions (ie, age, gender, duration of the disease and the use of ursodeoxycholic acid (UDCA)) were also requested. In addition, patients were asked to report possible presence of pruritus using the Visual Analogue Scale (VAS).[11] Participating members of the DLPA were asked for an additional written permission to contact their treating physician in order to confirm their diagnosis.

This study was divided into three phases, which consisted of (1) the development of the Dutch version of the PBC-40, (2) a pilot study testing the preliminary version of the Dutch PBC-40, and (3) the evaluation of the reliability and validity of the final version of the Dutch PBC-40 (and PBC-27) in a cohort of patients with PBC across the Netherlands.

Questionnaires

PBC-40

The PBC-40 is a HRQoL measure designed for self-completion in patients with PBC.[6] This disease-specific measure covers six domains through 40 quality of life statements, with higher scores denoting a worse HRQoL. The domain symptoms (7 items), itch (3 items), fatigue (11 items) and cognition (6 items) consist of a Likert scale of 1–5 points with 1 corresponding to 'never' and 5 corresponding to 'always'. Patients are requested to reflect on the past 4 weeks. The remaining domains, social (10 items) and emotional (3 items), do not refer to a specific time period and consist of a 1 to 5-point Likert scale with 1 corresponding to 'strongly disagree' and 5 corresponding to 'strongly agree'. For items 3, 8, 9, 10, 29 and 31, there is an additional 'does not apply' option available.

PBC-27

The PBC-27 is an abridged version of the PBC-40, which was developed in 2010 by Montali *et al*.[8] Contrary to the PBC-40, this scale redistributes its 27 items from six to seven domains: symptoms (3 items), dryness (2 items), itch (3 items), fatigue (8 items), cognition (5 items), emotional (3 items) and social (3 items). The scoring system does not differ from the original PBC-40.

RAND-36

The RAND-36 is the validated Dutch version of the 36-Item Short Form Health Survey (also known as SF-36) and is a widely used generic patient-reported outcome measure to assess HRQoL in various conditions.[12] It comprises eight health concepts (physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue and general health perceptions) which can be aggregated into two summary scores: 'mental component summary' and 'physical component summary'.[13] Scores on all individual items range from 0 to 100, with lower scores denoting a worse HRQoL.

Linguistic validation of the Dutch PBC-40 and pilot study

The PBC-40 was translated into Dutch following standardised forward–backward procedures.[7] The forward translation was independently performed by three native Dutch speakers with both medical and non-medical

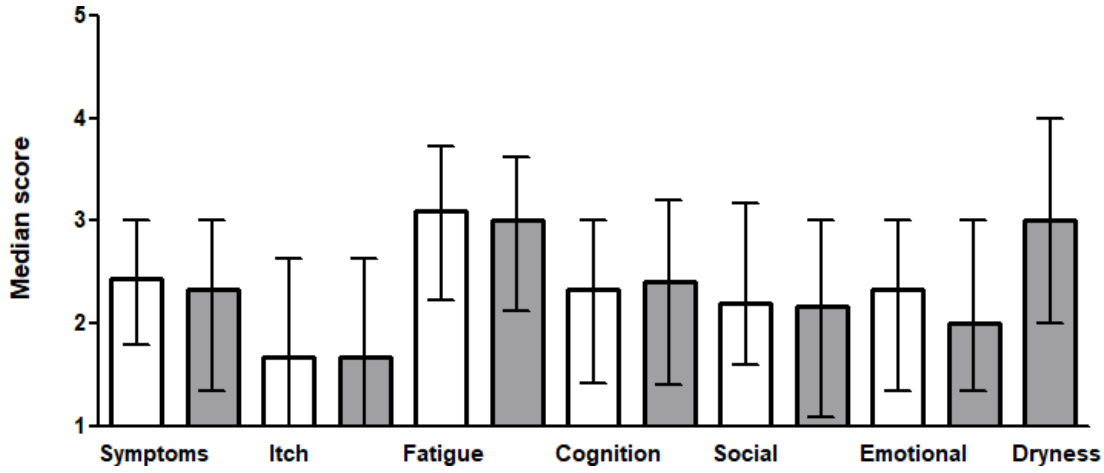

Figure 1 **Median domain scores for the PBC-40 and PBC-27.** The score of each domain ranges from 1 to 5. The white bars present the results of the PBC-40, and the grey bars present the results of the PBC-27. Error bars represent IQRs.

backgrounds. Discrepancies between the translations were discussed by the translators in order to reach consensus. This preliminary version of the Dutch PBC-40 was back-translated to English by a bilingual speaker (native English) with a medical background and then compared with the original version in order to identify misconceptions. As a pilot, three medical doctors without experience in the field of PBC and five patients with PBC were asked to complete this preliminary Dutch questionnaire. They were interviewed face-to-face by a member of the study team afterwards, during which the quality of life statements were evaluated per item. The feedback was discussed by the study team and minor textual edits were performed. This version of the Dutch PBC-40 was reviewed by PBC experts (who were part of the study team but were not involved in the back translation and forward translation). The finalised Dutch PBC-40 contains the same number of items, domains, layout and scoring system as the original PBC-40 (online supplemental material).

Statistical analyses

The measurement properties internal consistency, content validity and construct validity of the PBC-40 and PBC-27 were tested to assess the quality domains reliability and validity. If data were missing from a PBC-40 domain (defined as missed, duplicated or does not apply answers), the whole domain was discarded if <50% of items were completed. If >50% of the items were completed, the median value of the completed items in the domain was allocated to the missing items.[6]

The *internal consistency* is the degree of inter-relatedness among the items in the measure.[14] A reliable measure assesses a single underlying concept by using multiple items. This was calculated with Cronbach's alpha. A high Cronbach's alpha indicated high correlations between the multiple items. Values between 0.70 and 0.95 were considered to reflect adequate internal consistency.[15]

The *content validity* is the degree to which the content of a measure adequately reflects the target construct.[14] This was subjectively assessed and verified by examining whether the items appeared to be measuring what they are intended to measure ('face validity'); that is, the questions seemed plausible, relevant and to span the domain adequately. Face validity was determined by the entire study team (including the two PBC experts who were not involved in the back translation and forward translation), based on the linguistic validation process with feedback of three independent medical doctors and five patients with PBC, as described earlier. In addition, score distributions on floor and ceiling effects were analysed. These effects may occur when >15% of the patients score at the lower (floor) or upper (ceiling) end of the scale.

The *construct validity* is the extent to which scores relate to other known measures. Construct validity was verified by hypotheses testing using Pearson's correlation between RAND-36 scales and the various PBC-40 and PBC-27 domains (we expect negative Pearson's coefficients because the scales of the PBC-40/PBC-27 and the RAND-36 move in the opposite direction). The following hypotheses were formulated based on previous research[6 8]:

- PBC-40/PBC-27 domain symptoms negatively correlate with RAND-36 physical functioning and physical pain scales.
- PBC-40/PBC-27 domain itch negatively correlates with RAND-36 physical functioning scale and pruritus VAS score.
- PBC-40/PBC-27 domain fatigue negatively correlates with RAND-36 vitality scale.
- PBC-40/PBC-27 domain cognitive negatively correlates with RAND-36 mental component summary.
- PBC-40/PBC-27 domain social negatively correlates with RAND-36 social functioning scale.
- PBC-40/PBC-27 domain emotional negatively correlates with RAND-36 mental health and role emotional scale.
- PBC-27 domain dryness negatively correlates with RAND-36 physical functioning and physical role scale.

Table 2 Internal consistency of the Dutch PBC-40 and PBC-27

Domain	n	Cronbach's α PBC-40	Cronbach's α PBC-27
Symptoms	177	0.81	0.70
Dryness	177	–	0.69
Itch	111	0.83	0.83
Fatigue	177	0.95	0.92
Cognition	176	0.94	0.93
Social	176	0.91	0.85
Emotional	176	0.78	0.78

Construct validity was considered adequate when at least 75% of the aforementioned hypotheses were confirmed (ie, a statistically significant correlation).[14]

Data are presented as mean with SD in case of normal distribution, and otherwise as median with IQR. All statistical tests were two-sided, and a p value of <0.05 was considered to be statistically significant. Statistical analyses were performed using SPSS Statistics V.25.0.

RESULTS

Cohort characteristics

Following an overall response rate of 68%, a total of 177 patients with PBC were included. The mean age was 61.1 (SD 9.9) years, and the majority of the patients was female (n=164, 92.7%). More than half of the patients (n=97, 54.8%) reported a time since diagnosis of more than 10 years, and the large majority (n=166, 93.8%) were being treated with UDCA (table 1).

Table 3 Floor and ceiling effects of the Dutch PBC-40 and PBC-27

Domains (number of items)	Possible range	Actual range	n	Floor (%)	Ceiling (%)
PBC-40					
Symptoms (7)	7–35	7–32	176	1.7	0.0
Itch (3)	3–15	3–13	109	27.0	0.0
Fatigue (11)	11–55	11–53	177	1.7	0.0
Cognition (6)	6–30	6–30	176	19.3	0.6
Social (10)	10–50	10–49	176	5.1	0.0
Emotional (3)	3–15	3–15	176	10.8	1.7
PBC-27					
Symptoms (3)	3–15	3–15	177	11.9	0.6
Dryness (2)	2–10	2–10	177	6.8	7.3
Itch (3)	3–15	3–13	109	27.0	0.0
Fatigue (8)	8–40	8–38	177	2.8	0.0
Cognition (5)	5–25	5–25	176	19.3	0.6
Social (3)	3–15	3–15	176	25.0	2.3
Emotional (3)	3–15	3–15	176	10.8	1.1

Median scores of the different PBC-40 and PBC-27 domains are illustrated in figure 1. From the total of 7080 items, 61 items (0.9%) were missing (unanswered or duplicated answers) and 342 items (4.8%) were answered with the does not apply option.

Internal consistency

Each domain of the Dutch PBC-40 had a Cronbach's α coefficient above 0.70. The highest Cronbach's α coefficient was observed in the domain fatigue (0.95), followed by the domains cognition and social (0.94 and 0.91, respectively) (table 2). Comparable Cronbach's α coefficients were observed for the PBC-27 domains, except for the α coefficient of the domain symptoms, which was lower compared with the PBC-40 (0.70 vs 0.81). The PBC-27 specific domain dryness showed an α coefficient of 0.69.

Content validity

Face validity was considered adequate for the Dutch PBC-40 during the pilot study. Floor effects were present in two domains of the PBC-40 (cognition 19.3% and itch 27.0%) and in three domains of the PBC-27 (cognition 19.3%, social 25.0% and itch 27.0%). No ceiling effects were observed in both the PBC-40 and PBC-27 domain scores (table 3).

Construct validity

Pearson's correlations between RAND-36 scales and the various PBC-40 and PBC-27 domains are shown in table 4. Referring to the predefined hypotheses, all domains were statistically significantly correlated (p<0.001). The strongest correlation was observed between the PBC-40 domain fatigue and the vitality scale of the RAND-36 with a Pearson's correlation of −0.834 (p<0.001). Correlations between the PBC-27 domains and RAND-36 scales were comparable to the correlations between the PBC-40 and the RAND-36 scales (table 4). In addition, the itch domain of the PBC-40 and PBC-27 correlated significantly with the pruritus VAS score (r=−0.719, p<0.001).

DISCUSSION

We herein report the first validation study of our Dutch translation of the PBC-40 and PBC-27 which provides a useful PBC-specific HRQoL measure for Dutch-speaking patients with PBC. Internal consistency, reliability and validity were generally adequate. Based on our findings, the Dutch version of the PBC-40 and PBC-27 can be used in clinical practice, in HRQoL studies and in clinical drug trials.

Cronbach's alpha coefficients ranged from 0.81 to 0.95 for the different PBC-40 domains, which was largely similar to previous validation studies[8–10] and demonstrated adequate reliability. Interestingly, we observed a higher Cronbach's alpha for the domain symptoms (0.81) as compared with prior studies (range 0.52–0.72),[8–10] despite the heterogeneous nature of this domain. Montali *et al* developed the PBC-27, an abridged version of the

Table 4 Construct validity of PBC-40 and PBC-27

Domains	RAND-36 scales	Pearson's coefficient*	P value
PBC-40			
Symptoms	Physical functioning	−0.535	<0.001
Symptoms	Physical pain	−0.620	<0.001
Itch	Physical functioning	−0.322	<0.001
Fatigue	Vitality	−0.834	<0.001
Cognition	Mental component summary	−0.579	<0.001
Social	Social functioning	−0.779	<0.001
Emotional	Mental health	−0.708	<0.001
Emotional	Role emotional	−0.478	<0.001
PBC-27			
Symptoms	Physical functioning	−0.474	<0.001
Symptoms	Physical pain	−0.617	<0.001
Itch	Physical functioning	−0.322	<0.001
Dryness	Physical functioning	−0.423	<0.001
Dryness	Physical role	−0.262	<0.001
Fatigue	Vitality	−0.834	<0.001
Cognition	Mental component summary	−0.579	<0.001
Social	Social functioning	−0.744	<0.001
Emotional	Mental health	−0.646	<0.001
Emotional	Role emotional	−0.481	<0.001

*Pearson's coefficients are negative because the RAND-36 scales and PBC-40/PBC-27 scales move in the opposite direction. A better HRQoL translates to a higher RAND-36 score but lower PBC-40/PBC-27 scores.
HRQoL, health-related quality of life.

PBC-40, in which the items of the domain symptoms were redistributed into the domains symptom (three items) and dryness (two items) in order to improve the internal consistency. A known potential drawback of shortening questionnaires is the possibility of compromising its psychometric properties. Indeed, in our cohort, Cronbach's alpha of this PBC-27 domain symptoms declined from 0.81 to 0.70. However, the PBC-27 still demonstrated an adequate internal consistency in six out of the seven domains.

Adequate validity was demonstrated by the significant correlation between all PBC-40 and PBC-27 domains and their predefined specific RAND-36 scales. In addition, ceiling effects were entirely absent in both PBC-40 and PBC-27, indicating that both measures have the ability to determine HRQoL in patients with severe symptoms and therefore could be useful as an outcome measure in clinical trials. Floor effects were observed in the domains itch (27.0%) and cognition (19.3%), which were largely comparable to the observed floor effects in the original PBC-40 development study (36.7% and 12.5%, respectively).[6] The floor effect in the domain itch may be partly due to the ambiguous nature of the quality-of-life statements in this domain. Items 8, 9 and 10 refer to the impact of pruritus, and thus its severity. These items are meant to be answered only by those who are indeed experiencing pruritus. If pruritus is absent, patients are supposed to answer with the 'does not apply' option. However, 18.2% of patients who answered 'does not apply' still reported a pruritus VAS score of ≥1 in the previous 24 hours. It may be that these patients do experience pruritus, but that pruritus never disturbs sleep, makes them scratch their skin raw or makes them feel embarrassed. In this case, the answer 'never' for item 8, 9 and 10 would better reflect their clinical situation, which would increase the floor effect of these items in our study. A dichotomous question within this domain to indicate whether the patient ever experienced pruritus (yes or no) may be able to avoid potential misunderstanding and improve the registered responses. Remarkably, the floor effect in the domain social was 5.1% in the Dutch PBC-40 but increased to 25.0% in the Dutch PBC-27. This may be explained by the fact that the number of quality of life statements in this domain decreased from 10 to 3, whereby the chance of patients reporting the lowest possible score in this domain increased (resulting in an increased floor effect). This may indicate that the Dutch PBC-27 is less able to distinguish between patients with a low impact of PBC on their HRQoL in the social domain compared with the Dutch PBC-40. In addition, in line with the discussion on the statements related to itch above, the specific selection of the three statements which remain in the domain social of the PBC-27 may be of influence ('I can't go out and enjoy myself', 'I can't plan holidays' and 'My social life has stopped'). Other PBC-27 studies did not report floor and ceiling effects.

Jacoby *et al* reported the inability of patients to separate the effect of PBC on their quality of life from other (medical) conditions.[6] During our pilot study, we received similar feedback from our respondents. However, we decided not to make adjustments that could improve this issue because that could impact on patients' responses, which could hamper comparison of our results with the outcomes of previous research. For international scientific collaborations, uniform HRQoL measures are beneficial as well. Future studies could assess the impact of adjusted formulations, for instance, by removing 'because of PBC' from the statements. This may also prevent respondents from leaving items unanswered because they found the items to be irrelevant or difficult to answer.

Some limitations should be noted. First, the Dutch PBC-27 was not submitted separately to a new study population, which may be a potential limitation of this validation study. Although the psychometric properties of the

PBC-27 were not compromised by shortening the questionnaire, the PBC-40 did perform better in this validation study. However, the PBC-27 may be a valuable addition over the PBC-40 when a briefer measure is desired. Especially for physicians who are not familiar with the PBC-40, scoring 40 items can be complex and time consuming. Second, recruitment of patients through the DLPA may have led to a selection bias. Members of the DLPA may be more educated about their disease and aware of their symptoms. As a consequence, they may also be more inclined to participate in this study. However, even though the response rates were generally high, members of the DLPA responded less frequently as compared with patients of the Erasmus MC (62% vs 73%, respectively).

In conclusion, our findings demonstrate evidence for the appropriateness of our Dutch translation of both the PBC-40 and PBC-27 to measure HRQoL in patients with PBC in the Netherlands. As a result, we provide a useful PBC-specific HRQoL measure for Dutch-speaking patients with PBC that can be used for both research and clinical purposes.

Acknowledgements Thanks to Jacqueline Feyen and Mathijs R de Veer for their translations of the questionnaire, and especially to all the patients who gave their time and effort to participate.

Contributors RCdV and AJvdM had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of data analyses. Study concept and design: RCdV, GdS, EU and AJvdM. Acquisition of data: RCdV, GdS, MCvH and JW. Analysis and interpretation of data: RCdV, GdS, MCvH, MH, EU and AJvdM. Drafting of the manuscript: RCdV, GdS, EU and AJvdM. Critical revision of the manuscript for important intellectual content: all authors. Study supervision: HJM, EU and AJvdM. AJvdM is responsible for the overall content als guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests The following authors declared that they have no conflicts of interest: RCdV, GdS, MCvH, HJM, JW and EU. MH received a speakers fee from Zambon Nederland B.V. and thesis printing reimbursements from Intercept Pharmaceuticals, Pentax, Norgine, Tramedico, Dr. Falk, Sysmex, Astellas and Zambon Nederland B.V. AJvdM reports speakers fees from Zambon Nederland B.V., received an unrestricted grant from Gilead Sciences, AbbVie, MSD and Zambon Nederland B.V., and consulting work for AOP Pharma.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by ethics committee of the Erasmus Medical Center (reference number MEC-2020-0589) in Rotterdam and was conducted in accordance with the Declaration of Helsinki. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. Due to privacy and ethical concerns, data can not be shared.

ORCID iD

Rozanne C de Veer http://orcid.org/0000-0003-1604-0931

REFERENCES

1 Kaplan MM, Gershwin ME. Primary biliary cirrhosis. *N Engl J Med* 2005;353:1261–73.
2 Mells GF, Pells G, Newton JL, *et al*. Impact of primary biliary cirrhosis on perceived quality of life: the UK-PBC national study. *Hepatology* 2013;58:273–83.
3 Poupon RE, Chrétien Y, Chazouillères O, *et al*. Quality of life in patients with primary biliary cirrhosis. *Hepatology* 2004;40:489–94.
4 Carbone M, Mells GF, Pells G, *et al*. Sex and age are determinants of the clinical phenotype of primary biliary cirrhosis and response to ursodeoxycholic acid. *Gastroenterology* 2013;144:560–9.
5 European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu, European Association for the Study of the Liver. EASL clinical practice guidelines: the diagnosis and management of patients with primary biliary cholangitis. *J Hepatol* 2017;67:145–72.
6 Jacoby A, Rannard A, Buck D, *et al*. Development, validation, and evaluation of the PBC-40, a disease specific health related quality of life measure for primary biliary cirrhosis. *Gut* 2005;54:1622–9.
7 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417–32.
8 Montali L, Tanaka A, Riva P, *et al*. A short version of a HRQoL questionnaire for Italian and Japanese patients with primary biliary cirrhosis. *Dig Liver Dis* 2010;42:718–23.
9 Zec S, Popovic D, Matovic V, *et al*. Translation and validation of the Serbian primary biliary cholangitis-40 questionnaire. *PLoS One* 2017;12:e0175697.
10 Raszeja-Wyszomirska J, Wunsch E, Krawczyk M, *et al*. Assessment of health related quality of life in polish patients with primary biliary cirrhosis. *Clin Res Hepatol Gastroenterol* 2016;40:471–9.
11 Phan NQ, Blome C, Fritz F, *et al*. Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. *Acta Derm Venereol* 2012;92:502–7.
12 Van der Zee K, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36. Een handleiding. *Groningen* 2012.
13 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
14 Terwee CB, Bot SDM, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
15 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.