





A Systematic Review of Patient-Reported Outcomes in Primary Biliary Cholangitis and Primary Sclerosing Cholangitis

Hannah P. Kim ^{1*}, Sarah R. Lieber ^{1*}, Michael E. Rogers,² Andrew M. Moon ¹, Marci Loisel,³ Jennifer Walker,⁴ David N. Assis,⁵ Ricky Safer,⁶ Rachel Gomel,⁶ and Donna M. Evon ¹

Primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC) are associated with decreased health-related quality of life and debilitating symptoms. These experiences can be defined as patient-reported outcome (PRO) concepts and measured using PRO instruments. We identified all PRO concepts and instruments used in the PBC and PSC literature. This systematic review identified PBC and/or PSC studies from January 1, 1990, to May 6, 2019, that measured at least one PRO concept. Study population, design, PRO concept, PRO instrument, and validation data for PRO instruments were investigated. We provided descriptive statistics of PRO concepts and instruments used, stratified by population type. Use of PRO concepts and instruments were assessed over time. The search yielded 318 articles (69% in PBC, 18% in PSC, 13% in both, and 24% in drug trials). Forty-nine unique PRO concepts were identified. The five most common PRO concepts included pruritus (25%), fatigue (19%), broad health-related quality of life (16%), gastrointestinal adverse events (6%), and physical adverse events (6%). Only 60% of PRO concepts were measured with a PRO instrument, most of which were nonvalidated visual analogue or numeric rating scales. Only three of 83 PRO instruments were developed with feedback from the target populations (one for PBC, one for PSC, and one for both), and only six documented any psychometric testing in the target populations. Use of PRO instruments increased over time from 30% in the 1990s to 67% by 2019. *Conclusion:* The overwhelming majority of PRO instruments used in PBC/PSC were nonspecific and lacked patient validation or empirical justification. Significant opportunities exist to use qualitative methods to better understand patient experiences, and translate this knowledge into meaningful, patient-driven study outcomes. (*Hepatology Communications* 2020;4:1502-1515).

Primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC) are chronic cholestatic liver diseases associated with significant morbidity, including decreased quality of life,⁽¹⁻⁶⁾ pruritus,⁽⁷⁻⁹⁾ and fatigue,⁽¹⁰⁻¹⁴⁾ among other symptoms. The primary outcomes of most clinical trials include disease endpoints or surrogates (e.g., alkaline phosphatase), rather than patient-reported outcomes (PROs), which capture how patients feel or function. PROs reflect a patient's health status, as it comes directly from the patient, without interpretation by anyone else including the clinician.⁽¹⁵⁾ Patient experiences and perceptions are best evaluated using validated PRO instruments that were developed with

Abbreviations: FDA, Food and Drug Administration; HRQOL, health-related quality of life; NOS, not otherwise specified; PBC, primary biliary cholangitis; PRO, patient-reported outcome; PSC, primary sclerosing cholangitis.

Received April 7, 2020; accepted June 14, 2020.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1567/supinfo.

*These authors contributed equally to this work.

Supported by National Center for Advancing Translational Sciences (UL1TR002489), National Institutes of Health (T32 DK007634), and National Institute of Diabetes and Digestive and Kidney Diseases (P30-DK34987).

© 2020 The Authors. *Hepatology Communications* published by Wiley Periodicals LLC on behalf of the American Association for the Study of Liver Diseases. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

View this article online at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com).

DOI 10.1002/hep4.1567

Potential conflict of interest: Dr. Assis received grants from Gilead. Dr. Evon's institution received grants Gilead and Merck.

qualitative input from patients affected by the disease. Common PRO concepts evaluated in the biomedical literature include symptoms, health perceptions, quality of life, patient preferences, values related to decision making, and satisfaction with medical care.

In the last decade there has been burgeoning interest to use PRO instruments in clinical trials, to ensure that drug development embraces a more patient-centered approach. Recently, the US Food and Drug Administration (FDA) and other health care organizations have urged the use of PRO instruments to capture clinical outcomes, to collect meaningful patient input during drug clinical trials.^(15,16) In 2018, the FDA updated the guidelines for patient-focused drug development, to encourage the identification of concepts most important to patients, and to use PRO instruments that are validated in the target population to measure primary and secondary trial outcomes.⁽¹⁷⁾ PRO instruments are vital to capturing patient experiences and preferences, specifically in trials testing the safety and efficacy of medical therapies or interventions; this is vital to determining whether patients experience clinically meaningful improvements in areas of functioning that matter most to them. These clinical outcome assessments may, in turn, be used to seek FDA approval of new medical therapies.

To date, it remains unclear which PRO concepts (i.e., patient experiences or perceptions) and PRO instruments are most often used in the PBC and PSC scientific literature. The aims of this comprehensive systematic review were to identify (1) PRO concepts evaluated in the PBC and PSC scientific literature, (2) PRO instruments used to measure these PRO concepts, (3) types of studies using PRO instruments, and

(4) studies that describe the psychometric evaluation and validation of the PRO instruments being used.

Materials and Methods

SEARCH STRATEGY, DATABASES, KEY SEARCH TERMS

A comprehensive literature search was conducted in MEDLINE/PubMed, EMBASE and Scopus, to identify relevant articles from January 1, 1990, through May 6, 2019. The literature search included Medical Subject Headings, Emtree headings, and related text and keyword searches when appropriate, focusing on terms used to describe PROs in individuals with PBC or PSC (Supporting Table S1). The search strategy was developed with input from members of the research team and a librarian. To explore PRO concepts and measures in unpublished and ongoing clinical trials, additional searches were conducted in clinicaltrials.gov and clinicaltrialsregister.eu. The search terms “primary biliary cholangitis,” “primary biliary cirrhosis,” “PBC,” “primary sclerosing cholangitis,” and “PSC” were used to identify studies registered from January 1, 2013, through May 31, 2019, which were listed as completed, active, recruiting, or ongoing.

INCLUSION AND EXCLUSION SEARCH CRITERIA

Inclusion criteria were as follows: (1) Studies were conducted among individuals with a diagnosis of PBC or PSC; (2) a PRO concept or instrument

ARTICLE INFORMATION:

From the ¹Division of Gastroenterology and Hepatology, Department of Medicine, University of North Carolina, Chapel Hill, NC; ²Division of Pediatric Gastroenterology, Department of Pediatrics, University of North Carolina, Chapel Hill, NC; ³Department of Psychiatry and Behavioral Sciences, Division of Behavioral Medicine, Duke University Medical Center, Durham, NC; ⁴University of North Carolina, Health Sciences Library, Chapel Hill, NC; ⁵Digestive Diseases, Department of Medicine, Yale School of Medicine, New Haven, CT; ⁶PSC Partners Seeking a Cure, Greenwood Village, CO.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Donna M. Evon, Ph.D.
Division of Gastroenterology and Hepatology
University of North Carolina
Burnett-Womack Building CB #7584

Chapel Hill, NC 27599
E-mail: Donna_evon@med.unc.edu
Tel.: +(919) 966-6732

was included in the study methods or results; and (3) outcomes were reported specifically for patients with PBC or PSC. If more than one liver disease was studied, at least 50% of the study sample had to be diagnosed with PBC or PSC if no PRO instrument was used, or more than 2 patients had to be diagnosed with PBC or PSC if a PRO instrument was included. We excluded non-English studies, single-case reports, review articles, commentaries, editorials, and studies of patients with secondary or unspecified cholangitis.

For our clinical trial searches, we included studies if the sample included patients with PBC or PSC, if outcomes included a PRO concept or instrument, or if adverse events were reported in completed trials. Duplicate listings and terminated and published studies were excluded.

STUDY SELECTION AND DATA EXTRACTION

Six reviewers were trained and participated in the screening and data abstraction process. A minimum of two reviewers independently screened all titles and abstracts for inclusion using the eligibility criteria described previously. Studies with titles and abstracts that met inclusion criteria or lacked adequate information to determine inclusion or exclusion underwent full-text review. Conflicts were resolved by group consensus. During the full-text review, a minimum of two reviewers independently reviewed each full-text article for inclusion. If both reviewers agreed that a study did not meet eligibility criteria, the study was excluded. If the two reviewers disagreed, a third member of the review team was engaged to resolve the conflict. The following key variables were defined and extracted into a database: details of the study population, study design, year of publication, PRO concept, PRO instrument, and presence of any psychometric evaluation of the PRO measures.

Our study population was defined as including PBC, PSC, or both populations. It was also noted whether the study population focused on pediatric, pregnant, or post-liver transplant subpopulations. Studies were classified by design as follows: drug randomized controlled trials; drug-related retrospective, prospective, cross-sectional or case series; non-drug retrospective, prospective, cross-sectional or case series; development or validation of PRO instrument studies; qualitative;

mixed methods; cost-effectiveness or health utilities; and surgical or procedural intervention studies. PRO concepts were classified as to whether they referred to a (1) study outcome that was clearly defined in the methods as an *a priori* outcome or (2) descriptive variable that was limited to a PRO concept to describe the study population (e.g., in Table 1 of a study) or an adverse effect. PRO concepts were also classified into four categories: (1) quality of life/functioning, (2) symptoms, (3) adverse events or side effects, and (4) other. Adverse events included side effects from medications or a study intervention that were classified as cognitive (e.g., memory impairment), emotional (e.g., depression), gastrointestinal (e.g., abdominal pain), or physical (e.g., pruritus).

TABLE 1. CHARACTERISTICS OF PBC AND PSC STUDIES INCLUDED IN THE SYSTEMATIC REVIEW (n = 318)

Variable	Frequency, n (%)
Patient population	
PBC	220 (69.2)
PSC	56 (17.6)
Both	42 (13.2)
Special populations	
Pediatric	4 (1.3)
Pregnant	5 (1.6)
Transplant	18 (5.7)
Trial design/type of intervention investigated	
Drug RCT	77 (24.2)
Drug prospective	40 (12.6)
Drug retrospective	12 (3.8)
Drug case series	7 (2.2)
Non-drug cross-sectional	72 (22.7)
Non-drug prospective	34 (10.7)
Non-drug retrospective	30 (9.4)
Non-drug case series	8 (2.5)
PRO measure development or validation	11 (3.5)
Qualitative only	3 (0.9)
Mixed methods	3 (0.9)
Surgical/procedural	19 (6.0)
Cost-effectiveness/health utilities	2 (0.6)
Years	
1990-1994	47 (14.8)
1995-1999	38 (12.0)
2000-2004	42 (13.2)
2005-2009	59 (18.5)
2010-2014	60 (18.9)
2015-2019	72 (22.6)

Abbreviation: RCT, randomized controlled trial.

DATA ANALYSIS

We summed the total number of times a PRO concept was used for each study population. For example, if one study evaluated the PRO concepts of health-related quality of life (HRQOL), fatigue, and pain (concepts = 3), and a second study evaluated the PRO concepts of HRQOL, fatigue, nausea, and anxiety (concepts = 4), a total sum of seven concepts were evaluated between the two studies. We also calculated the number of *unique* PRO concepts across all studies (e.g., five unique PRO concepts between the two studies mentioned previously). Descriptive statistics for categorical variables were provided and stratified by population (PBC, PSC, or both). We analyzed data on the individual study level and at the PRO concept level. Use of PRO concepts and instruments was also assessed over time. All analyses were performed using STATA version 15.0 (StataCorp LP, College Station, TX).

Results

SEARCH RESULTS

A total of 6,773 articles were identified through database searching, of which 4,833 were nonduplicates (Fig. 1). A total of 4,424 articles were excluded during abstract screening, leaving 409 articles eligible for full-text review. An additional 91 articles were excluded

during the full-text review phase. A total of 318 articles met all eligibility criteria and were included in the study (Supporting Table S2). A total of 152 clinical trial listings were identified in clinicaltrials.gov and clinicaltrialsregister.eu, of which 32 were included in this review after excluding trials that were duplicated, terminated, or published.

STUDY POPULATION: ARTICLE TYPE AND PARTICIPANT POPULATION

Almost a quarter of the studies were published after 2015, indicating a substantial increase in the literature's focus on PRO concepts in PBC and PSC (Table 1). The largest number of studies describing PRO concepts were drug trials (24%) and non-drug cross-sectional studies (23%). Most studies were conducted in the population of patients with PBC (69%), while only 18% involved patients with PSC and 13% included both disease populations. Very few studies specified involvement of pediatric (1%), pregnant (2%), or liver transplant (6%) subpopulations.

OVERALL PRO CONCEPTS AND INSTRUMENTS IN THE PBC AND PSC LITERATURE

Among the 318 PBC or PSC studies, a sum total of 926 PRO concepts were evaluated or described

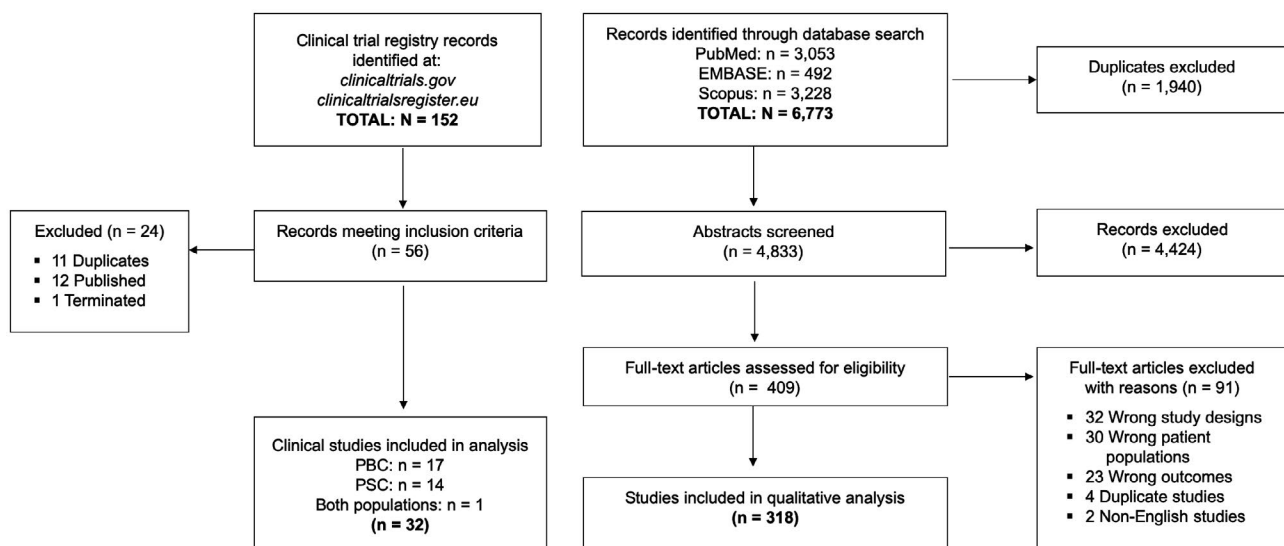


FIG. 1. Flowchart summarizing study identification and selection.

that could be captured directly by patient self-report. Of this total, we identified 49 unique PRO concepts (Table 2). In total, 83 PRO instruments were used to measure the 49 unique PRO concepts in the literature (Supporting Table S3). Most of the PRO concepts reflected study outcomes, while 27% were simply used to describe characteristics of the study population (e.g., proportion of patients with baseline pruritus). The top five most commonly cited PRO concepts included pruritus/itch (25%), fatigue (19%), broad HRQOL (16%), gastrointestinal side effects (6%), and physical side effects (6%). Overall, only 60% of PRO concepts were evaluated using a PRO instrument (Tables 3 and 4). Furthermore, when a PRO instrument was used, only 21% of these studies provided any information on the development or psychometric properties of the instrument itself, and most studies did not use validated instruments. Among the 40% of studies that did not use a PRO instrument, data on the PRO concept were extracted from medical records, clinician-rated, or not specified. Given that patient-reported adverse events (AEs) are not typically assessed using PRO instruments, we performed a sensitivity analysis excluding PRO concepts that were adverse events or side effects. When AEs were removed from the analysis, there was an increase in the proportion of PRO concepts measured by an instrument from 60% to 68%; however, the use of validated instruments remained infrequent.

To investigate the evolution of PRO concept and instrument use, we stratified publications over time (Fig. 2). The proportion of PRO concepts evaluated using an instrument increased from 30% in the early 1990s to 45% in the early 2000s and 67% by 2019. The most pronounced increase occurred between 2000 and 2010. However, the use of PRO instruments lagged behind the investigation of PRO concepts, which remained primarily captured through medical record review, clinician ratings, or clinician grading systems.

PRO CONCEPTS AND INSTRUMENTS IN PBC

A total of 642 PRO concepts were referred to in PBC studies, including 47 unique PRO concepts (Table 2). Most of the PRO concepts were related to specific symptoms (66%), whereas fewer concepts reflected HRQOL (16%) or adverse side effects (14%). The five most commonly evaluated symptoms were

pruritus/itch (26%), fatigue (20%), depression (4%), sleepiness/daytime somnolence (3%), and anxiety (2%). Broad HRQOL was evaluated in 13% of studies and addressed multiple domains of disease impact and functioning. In drug trials, 23 unique PRO concepts were evaluated primarily related to AEs, including gastrointestinal, physical, cognitive, and emotional adverse events.

Approximately 60% of the time, a clearly identified instrument was used to measure a PRO concept in patients with PBC (Table 3). The most commonly used type of PRO instrument was an unlabeled, non-specific version of a numeric rating or Likert scale (10%). The PBC-40 was used in 9% of studies, followed by the Fatigue Impact Scale (6%) and an itch visual analog scale (6%). The PRO instruments that were used to evaluate the five most common PRO concepts are depicted in Table 4. The concept of pruritus was measured using a PRO instrument only about half of the time (54%). Over 10 different PRO instruments were used to measure pruritus, and only two of these underwent appropriate validation and psychometric testing in the PBC population (PBC-40, Itch Diary).^(18,19) Fatigue was assessed using a PRO instrument about 60% of the time, and was measured using over 10 different instruments, only one of which was developed with input from patients with PBC (PBC-40). Two other instruments (Fatigue Impact Scale and Fisk Fatigue Severity Scale) were developed in other patient populations but subsequently underwent psychometric testing in the PBC population.^(20,21) In contrast, broad HRQOL was more often assessed using a validated PRO instrument, namely the PBC-40 (43%), Chronic Liver Disease Questionnaire (CLDQ) (6%), or the PBC-27 (5%).^(18,22,23) However, 46% of the time, HRQOL was measured using instruments not developed or validated in PBC, most commonly the SF-36 (24%). Over 90% of gastrointestinal and physical AEs that were reported were not measured using a PRO instrument. PRO instruments used to assess AEs included unspecified Likert or numeric rating scales, or diaries. No validated instruments were used to assess AEs in PBC trials.

PRO CONCEPTS AND INSTRUMENTS IN PSC

A total of 177 PRO concepts were cited in the PSC literature, including 47 unique PRO concepts

TABLE 2. NUMBER OF TIMES A PRO CONCEPT WAS MENTIONED IN THE PBC AND PSC LITERATURE (n = 926)

Variable	Frequency n (%)		
	PBC (n = 642)	PSC (n = 177)	Both (n = 107)
Concept category			
Quality of life/functioning	105 (16.4)	38 (21.5)	32 (29.9)
Symptoms	421 (65.6)	120 (67.8)	51 (47.7)
Treatment AEs/side effects	90 (14.0)	18 (10.2)	22 (20.5)
Other	26 (4.0)	1 (0.5)	2 (1.9)
Descriptive vs. outcome			
Descriptive only	174 (27.1)	48 (27.1)	25 (23.4)
Outcome only	468 (72.9)	129 (72.9)	82 (76.6)
Specific concepts			
Quality of life/functioning			
Broad HRQOL	82 (12.8)	34 (19.2)	30 (28.0)
Cognitive functioning	4 (0.6)	0 (0)	0 (0)
Emotional functioning	3 (0.5)	1 (0.6)	2 (1.9)
Physical functioning	8 (1.3)	1 (0.6)	1 (0.9)
Social functioning	4 (0.6)	0 (0)	1 (0.9)
Symptoms			
Abdominal distension	2 (0.3)	0 (0)	0 (0)
Abdominal pain	14 (2.2)	23 (13)	0 (0)
Anorexia	6 (0.9)	2 (1.1)	0 (0)
Anxiety	15 (2.3)	0 (0)	0 (0)
Arthralgia	6 (0.9)	1 (0.6)	0 (0)
Autonomic dysfunction	8 (1.3)	1 (0.6)	2 (1.9)
Back pain	1 (0.2)	0 (0)	0 (0)
Blood in stool	1 (0.2)	1 (0.6)	0 (0)
Bone pain	2 (0.3)	0 (0)	0 (0)
Depression	25 (3.9)	4 (2.3)	5 (4.7)
Diarrhea	3 (0.5)	5 (2.8)	0 (0)
Edema	1 (0.2)	0 (0)	0 (0)
Fatigue	130 (20.3)	28 (15.8)	14 (13.1)
Fever	0 (0)	2 (1.1)	0 (0)
Malaise	2 (0.3)	2 (1.1)	0 (0)
Muscle aches	2 (0.3)	0 (0)	0 (0)
Nausea/vomiting	4 (0.6)	5 (2.8)	0 (0)
Other gastrointestinal symptoms (multiple)	1 (0.2)	1 (0.6)	0 (0)
Poor sleep quality	1 (0.2)	0 (0)	2 (1.9)
Pruritus/itch	166 (25.9)	39 (22.0)	26 (24.3)
Sexual dysfunction	1 (0.2)	2 (1.1)	0 (0)
Sicca symptoms	1 (0.2)	0 (0)	0 (0)
Sleep disturbance	12 (1.9)	1 (0.6)	0 (0)
Sleepiness/daytime somnolence	17 (2.7)	1 (0.6)	0 (0)
Weight loss	0 (0)	3 (1.7)	0 (0)
Xerophthalmia	4 (0.6)	0 (0)	0 (0)
Xerostomia	3 (0.5)	0 (0)	0 (0)
Treatment AEs/side effects			
Cognitive AEs	6 (0.9)	0 (0)	2 (1.9)
Emotional AEs	2 (0.3)	1 (0.9)	0 (0)

TABLE 2. *Continued*

Variable	Frequency n (%)		
	PBC (n = 642)	PSC (n = 177)	Both (n = 107)
Gastrointestinal AEs	41 (6.4)	10 (5.7)	8 (7.5)
Opioid withdrawal symptoms	2 (0.3)	0 (0)	2 (1.9)
Physical AEs	34 (5.3)	8 (4.5)	9 (8.4)
Other			
Coping strategies	2 (0.3)	0 (0)	0 (0)
Dissatisfaction with health care delivery or information	1 (0.2)	0 (0)	0 (0)
Health behaviors (e.g., smoking, exercise)	1 (0.2)	0 (0)	1 (0.9)
Health perception	2 (0.3)	0 (0)	1 (0.9)
Identity/role change	1 (0.2)	0 (0)	0 (0)
Illness perceptions	2 (0.3)	0 (0)	0 (0)
Loneliness/social isolation	2 (0.3)	0 (0)	0 (0)
Psychological distress	4 (0.6)	1 (0.6)	0 (0)
Self-efficacy	3 (0.5)	0 (0)	0 (0)
Social support	1 (0.2)	0 (0)	0 (0)
Stigma	6 (0.9)	0 (0)	0 (0)
Uncertainty	3 (0.5)	1 (0.6)	0 (0)

(Table 2). Most of the PRO concepts reflected disease symptoms (68%), and fewer concepts related to functioning and HRQOL (22%), or treatment-adverse events (10%). The five most commonly reported symptoms were pruritus (22%), fatigue (16%), abdominal pain (13%), diarrhea (3%), and nausea/vomiting (3%). The broad concept of HRQOL was measured in 19% of studies. Among drug trials, 11 unique PRO concepts were explored among 53 references to PRO concepts.

Within the PSC literature, only half of the PRO constructs mentioned were clearly measured using a PRO instrument (Table 3). Similar to in PBC, the most commonly used PRO instrument was an unspecified Likert or numeric rating scale. The next most commonly used PRO instruments were unspecified visual analog scales (7%), the SF-36 (6%), as well as the PBC-40, Itch Visual Analog Scale, and Fatigue Impact Scale, each of which were used 3% of the time. Pruritus and fatigue were measured using a PRO instrument approximately half of the time; the remainder of the time, it is assumed that symptom data were retrieved from medical records or graded by clinicians (Table 4). Six unique instruments were used to measure pruritus, of which none were developed or validated in the PSC population. Of the four unique instruments that assessed fatigue, only the PBC-40 was validated in the PSC population.⁽²⁴⁾ Broad HRQOL was assessed using a PRO instrument 97%

of the time, most often with the generic SF-36 (32%), PBC-40 (15%), and the CLDQ (12%). The CLDQ was developed with input from patients with PSC, and the PBC-40 and PBC-27 underwent psychometric testing among patients with PSC.^(22,24) The new PSC-PRO was only recently developed in accordance with FDA guidelines, including published evidence of content validation and psychometric testing.⁽²⁵⁾ Gastrointestinal and physical AEs were never measured using a PRO instrument.

SPECIAL SUBPOPULATIONS

Of the 318 studies, four included a pediatric population and evaluated seven unique PRO concepts, five included pregnant women and evaluated three PRO concepts, and 18 included post-liver transplant patients and evaluated 17 unique PRO concepts (Supporting Table S2). Among studies including pediatric patients, pruritus was evaluated in all four studies, and abdominal pain was evaluated in three studies. Pruritus was evaluated in all five studies, including pregnant patients. Most of the PRO concepts were not measured using a PRO instrument or were measured using a nonvalidated or disease-nonspecific instrument. Among studies including posttransplant patients, broad HRQOL was the most common concept evaluated in all 18 studies and quantified using 9 different PRO instruments among 17 studies.

TABLE 3. PRO INSTRUMENTS USED TO MEASURE PRO CONCEPTS IN THE PBC AND PSC LITERATURE

Variable	Frequency n (%)			
	PBC (n = 642)	PSC (n = 177)	Both (n = 107)	Total (n = 926)
Use of PRO Instrument				
Yes	381 (59.3)	93 (52.5)	84 (78.5)	558 (60.2)
No	242 (37.7)	68 (38.4)	23 (21.5)	333 (36.0)
Unclear	19 (3.0)	16 (9.1)	0 (0)	35 (3.8)
Mention of PRO instrument validity in PBC/PSC				
Yes	87 (13.6)	14 (7.9)	14 (13.1)	115 (12.4)
No	294 (45.8)	79 (44.6)	70 (65.4)	443 (47.8)
Not applicable (no instrument)	261 (40.6)	84 (47.5)	23 (21.5)	368 (39.4)
TOP 20 most commonly used PRO instruments				
<i>Likert/Numeric Grading Scale NOS</i>	63 (9.8)	18 (10.2)	3 (2.8)	84 (9.1)
PBC-40	58 (9.0)*	6 (3.4) [†]	4 (3.7)	67 (7.2)
<i>Itch Visual Analog Scale</i>	35 (5.5)	6 (3.4)	18 (16.8)	59 (6.4)
Fatigue Impact Scale	36 (5.6) [†]	6 (3.4)	4 (3.7)	46 (5.0)
<i>SF-36</i>	20 (3.1)	11 (6.2)	13 (12.2)	44 (4.8)
<i>Visual Analog Scale NOS</i>	9 (1.4)	12 (6.8)	9 (8.4)	30 (3.2)
<i>Hospital Anxiety and Depression Scale</i>	23 (3.6)	2 (1.1)	1 (0.9)	26 (2.8)
<i>Epworth Sleepiness Scale</i>	15 (2.3)	1 (0.6)	0 (0)	16 (1.7)
Chronic Liver Disease Questionnaire	5 (0.8)*	4 (2.3)*	3 (2.8)	12 (1.3)
<i>5-D Itch Scale</i>	9 (1.4)	2 (1.1)	0 (0)	11 (1.2)
<i>Diary NOS</i>	7 (1.1)	4 (2.3)	0 (0)	11 (1.2)
<i>Orthostatic Grading Scale—Autonomic Events</i>	7 (1.1)	0 (0)	2 (1.9)	9 (1.0)
PBC-27	4 (0.6)*	3 (1.7) [†]	2 (1.9)	9 (1.0)
Fisk Fatigue Severity Scale	5 (0.8) [†]	0 (0)	3 (2.8)	8 (0.9)
<i>Beck Depression Inventory</i>	4 (0.6)	1 (0.6)	1 (0.9)	6 (0.7)
<i>Pittsburgh Sleep Quality Index</i>	6 (0.9)	0 (0)	0 (0)	6 (0.7)
<i>Pruritus Numeric Rating Scale (0-10)</i>	2 (0.3)	2 (1.1)	1 (0.9)	5 (0.5)
<i>Gastrointestinal Symptom Rating Scale</i>	3 (0.5)	1 (0.6)	0 (0)	4 (0.4)
<i>Hamilton Depression Rating Scale</i>	3 (0.5)	0 (0)	1 (0.9)	4 (0.4)
<i>Krupp's Fatigue Severity Scale</i>	4 (0.6)	0 (0)	0 (0)	4 (0.4)

Note: Bolded instruments denote instruments that were either developed in the target population or subsequently underwent psychometric testing in the target population.

*Instrument was developed with input from the target population.

[†]Instrument was not developed with input from the target population but later underwent psychometric testing in the target population.

Abbreviation: NOS, not otherwise specified.

CLINICAL REGISTRIES OF PBC AND PSC TRIALS

We identified 32 PBC or PSC ongoing or completed studies in the United States and European Union clinical trial registries from 2012 to 2019 that mention a PRO concept. Among these, 53% are in the PBC population, 44% are in the PSC population, and 3% include both populations. Over half of the studies (56%) investigate drug therapeutics in a randomized controlled trial.

Overall, PRO concepts were mentioned in these studies 88 times; 77 pertained to quality of life/functioning, symptoms, or health perceptions; and 11 represented treatment-adverse events (Table 5). Of the 11 AE PRO concepts mentioned, only three were mentioned as *a priori* study outcomes, and only one of these outcomes will be measured using a PRO instrument. The 77 non-AE PRO concepts represented 14 unique concepts; the most common concepts listed as study outcomes included pruritus (33%), broad HRQOL (27%), and fatigue (11%).

TABLE 4. PRO INSTRUMENTS USED TO MEASURE THE TOP 5 MOST COMMON PRO CONCEPTS IN THE PBC AND PSC LITERATURE

PRO Concept/Instrument	Frequency, n (%)			
	PBC (n = 642)	PSC (n = 177)	Both (n = 107)	Total Concepts (n = 926)
PRURITUS TOTAL	166 (25.9)	39 (22.0)	26 (24.3)	231 (25.0)
<i>No Instrument</i>	76 (45.8)	22 (56.4)	1 (3.8)	99 (42.9)
<i>Itch Assessment With Visual Analog Scale</i>	33 (19.9)	6 (15.4)	18 (69.2)	57 (24.7)
<i>Likert/Numeric Grading Scale NOS</i>	28 (16.9)	6 (15.4)	2 (7.7)	36 (15.6)
<i>5-D Itch Scale</i>	9 (5.4)	1 (2.6)	0 (0)	10 (4.3)
PBC-40	8 (4.8)*	0 (0)	1 (3.8)	9 (3.9)
<i>Pruritus numeric rating scale (0-10)</i>	2 (1.2)	2 (5.1)	1 (3.8)	5 (2.2)
<i>Diary NOS</i>	2 (1.2)	1 (2.6)	0 (0)	3 (1.3)
<i>Visual Analog Scale NOS</i>	2 (1.2)	1 (2.6)	0 (0)	3 (1.3)
<i>Itch Diary (Patient-Reported Symptom Questionnaire)</i>	1 (0.6)*	0 (0)	1 (3.8)	2 (0.9)
<i>Itch Severity Scale</i>	1 (0.6)	0 (0)	1 (3.8)	2 (0.9)
<i>Other[‡]</i>	4 (2.4)	0 (0)	1 (3.8)	5 (2.0)
FATIGUE TOTAL	130 (20.3)	28 (15.8)	14 (13.1)	172 (18.6)
<i>No instrument</i>	50 (38.5)	14 (50.0)	0 (0)	64 (37.2)
<i>Fatigue Impact Scale</i>	35 (26.9) [†]	6 (21.4)	4 (28.6)	45 (26.2)
<i>Likert/Numeric Grading Scale NOS</i>	12 (9.2)	3 (10.7)	0 (0)	15 (8.7)
PBC-40	13 (10.0)*	1 (3.6) [†]	0 (0)	14 (8.1)
<i>Visual Analog Scale NOS</i>	4 (3.1)	4 (14.3)	6 (42.9)	14 (8.1)
<i>Fisk Fatigue Severity Scale</i>	5 (3.9) [†]	0 (0)	3 (21.4)	8 (4.7)
<i>Krupp's Fatigue Severity Scale</i>	4 (3.1)	0 (0)	0 (0)	4 (2.3)
<i>Fatigue Diary</i>	3 (2.3)	0 (0)	0 (0)	3 (1.7)
<i>Modified Fatigue Impact Scale</i>	2 (1.5)	0 (0)	1 (7.1)	2 (1.2)
<i>Other[‡]</i>	2 (1.4)	0 (0)	0 (0)	3 (1.8)
BROAD HRQOL TOTAL	82 (12.8)	34 (19.2)	30 (28.0)	146 (15.8)
<i>SF-36</i>	20 (24.4)	11 (32.4)	13 (43.3)	44 (30.1)
PBC-40	35 (42.7)*	5 (14.7) [†]	3 (10.0)	42 (28.8)
<i>Chronic Liver Disease Questionnaire</i>	5 (6.1)*	4 (11.8)*	3 (10.0)	12 (8.2)
PBC-27	4 (4.9)*	3 (8.8) [†]	1 (3.3)	8 (5.5)
<i>Likert/Numeric Grading Scale NOS</i>	4 (4.9)	1 (2.9)	0 (0)	5 (3.4)
<i>Nottingham Health Profile</i>	3 (3.7)	0 (0)	0 (0)	3 (2.1)
<i>NIDDK-QA</i>	1 (1.2)	0 (0)	2 (6.7)	3 (2.1)
<i>Visual Analog Scale NOS</i>	3 (3.7)	0 (0)	0 (0)	3 (2.1)
<i>15D</i>	0 (0)	1 (2.9)	1 (3.3)	2 (1.4)
<i>No instrument</i>	1 (1.2)	1 (2.9)	0 (0)	2 (1.4)
<i>Other[‡]</i>	7 (8.5)	8 (23.5)	7 (23.3)	22 (15.1)
GASTROINTESTINAL AEs/TREATMENT SIDE EFFECTS TOTAL	41 (6.4)	10 (5.7)	8 (7.5)	59 (6.4)
<i>No instrument</i>	37 (90.2)	10 (100)	8 (100)	55 (93.2)
<i>Likert/Numeric Grading Scale NOS</i>	2 (4.9)	0 (0)	0 (0)	2 (3.4)
<i>Diary NOS</i>	1 (2.4)	0 (0)	0 (0)	1 (1.7)
<i>Gastrointestinal Symptom Rating Scale</i>	1 (2.4)	0 (0)	0 (0)	1 (1.7)
PHYSICAL AEs/TREATMENT SIDE EFFECTS TOTAL	34 (5.3)	8 (4.5)	9 (8.4)	51 (5.5)
<i>No instrument</i>	31 (91.2)	7 (87.5)	9 (100)	47 (92.1)
<i>Likert/Numeric Grading Scale NOS</i>	1 (2.9)	1 (12.5)	0 (0)	2 (3.9)

TABLE 4. Continued

PRO Concept/Instrument	Frequency, n (%)			
	PBC (n = 642)	PSC (n = 177)	Both (n = 107)	Total Concepts (n = 926)
<i>Diary NOS</i>	1 (2.9)	0 (0)	0 (0)	1 (2.0)
<i>Itch Assessment With Visual Analog Scale</i>	1 (2.9)	0 (0)	0 (0)	1 (2.0)

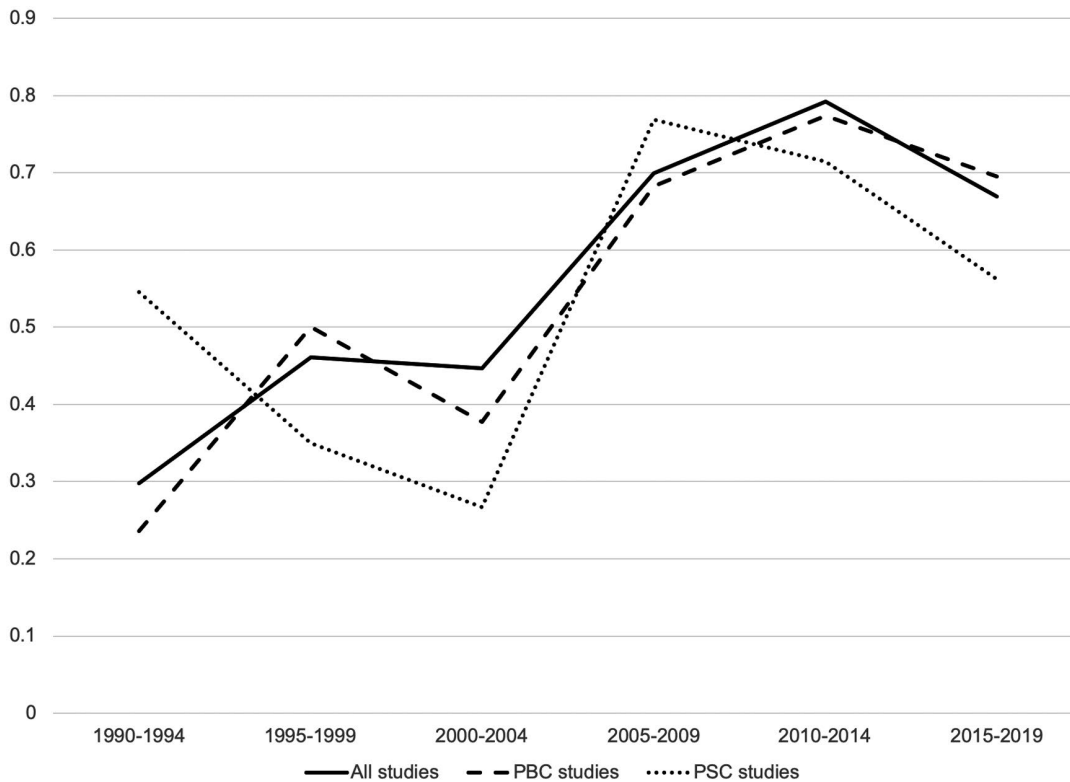
Note: Bolded instruments denote instruments that were either developed in the target population or subsequently underwent psychometric testing in the target population.

*Instrument was developed with input from the target population.

†Instrument was not developed with input from the target population but later underwent psychometric testing in the target population.

‡Pooled total of remainder of instruments used to measure the PRO concept (given the low frequency of each instrument being used per PRO concept found in the total population).

Abbreviation: GI, gastrointestinal.



HEP4-20-0101

FIG. 2. Proportion of PRO concepts measured using PRO instruments over time in the published literature of populations with PBC and PSC: 1990-2019.

PRO instruments were used to assess 96% of pre-specified study outcomes (Table 6). While almost all of these *a priori* study outcomes were measured using PRO instruments, the choice of instrument varied widely, with most instruments having no empirical documentation of being developed or validated in the target population.

Discussion

This comprehensive systematic review identified PRO concepts and instruments described in PBC and PSC studies over the last 30 years. Despite many studies involving a PRO concept, PRO instruments were used to measure these concepts only 50%-60%

TABLE 5. NUMBER OF TIMES A PRO CONCEPT IS MENTIONED IN UNPUBLISHED AND ONGOING STUDIES LISTED IN US AND EU CLINICAL TRIALS REGISTRIES FOR PBC AND PSC

PRO Concepts	Frequency, n (%)		
	PBC (n = 50)	PSC (n = 34)	Total (n = 88)
Functioning/quality of life			
<i>Broad HRQOL</i>	14 (28.0)	9 (26.5)	24 (27.3)
<i>Cognitive functioning</i>	1 (2.0)	0 (0.0)	1 (1.1)
<i>Physical functioning</i>	1 (2.0)	1 (2.9)	2 (2.3)
Symptoms			
<i>Anxiety</i>	1 (2.0)	0 (0.0)	1 (1.1)
<i>Autonomic dysfunction</i>	1 (2.0)	0 (0.0)	1 (1.1)
<i>Depression</i>	1 (2.0)	0 (0.0)	1 (1.1)
<i>Fatigue</i>	5 (10.0)	4 (11.8)	10 (11.4)
<i>Multiple gastrointestinal symptoms</i>	0 (0.0)	1 (2.9)	1 (1.1)
<i>Poor sleep quality</i>	2 (4.0)	0 (0.0)	2 (2.3)
<i>Pruritus/itch</i>	19 (38.0)	8 (23.5)	29 (33.0)
<i>Sleep disturbance</i>	0 (0.0)	1 (2.9)	1 (1.1)
<i>Sleepiness/daytime somnolence</i>	1 (2.0)	0 (0.0)	1 (1.1)
<i>Weakness</i>	1 (2.0)	0 (0.0)	1 (1.1)
Treatment AEs/side effects			
<i>Gastrointestinal AEs</i>	2 (4.0)	4 (11.8)	6 (6.8)
<i>Physical AEs</i>	1 (2.0)	4 (11.8)	5 (5.7)
Other			
<i>Health perception</i>	0 (0.0)	2 (5.9)	2 (2.3)

of the time. We observed tremendous variation in the types of instruments being used to evaluate the most common PRO concepts of pruritus, fatigue, and broad HRQOL. Moreover, most instruments were not developed with qualitative input from patients living with PBC or PSC, raising the question of whether these items appropriately capture the experiences and perceptions of these patients. This uncertainty likely exists, in part, due to a woefully low number of qualitative studies (e.g., in-depth interviews, focus groups) with patients with PBC or PSC, which stymies our understanding of their subjective experiences and perspectives. These findings highlight significant opportunities to improve the ways we capture disease symptoms, treatment side effects, and patient preferences using validated PRO instruments for PBC and PSC.

In the field of PBC, the PBC-40, published in 2005, is the only PRO instrument that included formative patient interviews to guide the development of the measure, and thus has strong content validity

TABLE 6. PRO INSTRUMENTS USED TO MEASURE THE TOP 3 MOST COMMON PRO CONCEPTS IN UNPUBLISHED AND ONGOING STUDIES FOR PBC AND PSC LISTED IN US AND EU CLINICAL TRIALS REGISTRIES

PRO Instruments	Frequency, n (%)		
	PBC	PSC	Total
PRURITUS TOTAL	19 (100)	8 (100)	29 (100)
<i>No instrument</i>	1 (5.3)	1 (12.5)	2 (6)
<i>Itch Visual Analog Scale</i>	7 (36.8)	6 (75.0)	14 (48.3)
<i>5-D Itch Scale</i>	5 (26.3)	—	6 (20.7)
<i>Pruritus Numeric Rating Scale (0-10)</i>	3 (15.8)	—	3 (10.3)
<i>ItchRo E-diary</i>	1 (5.3)	1 (12.5)	2 (6.9)
PBC-40	1 (5.3)*	—	1 (3.5)
<i>Questionnaire NOS</i>	1 (5.3)	—	1 (3.5)
FATIGUE TOTAL	5 (100)	4 (100)	10 (100)
<i>No instrument</i>	—	—	—
<i>Krupp's Fatigue Severity Scale</i>	1 (20.0)	2 (50.0)	3 (30.0)
<i>Likert/Numeric Grading Scale NOS</i>	2 (40.0)	—	2 (20.0)
<i>Fatigue Diary</i>	1 (20.0)	—	1 (10.0)
<i>Modified Fatigue Impact Scale</i>	—	1 (25.0)	1 (10.0)
PBC-40	1 (20.0)*	—	1 (10)
<i>Questionnaire NOS</i>	—	—	1 (10.0)
<i>Visual Analog Scale NOS</i>	—	1 (25.0)	1 (10)
BROAD HRQOL TOTAL	14 (100)	9 (100)	24 (100)
<i>No instrument</i>	—	1 (11.1)	1 (4.2)
PBC-40	8 (57.1)*	1 (11.1)†	9 (37.5)
<i>SF-36</i>	3 (21.4)	2 (22.2)	5 (20.8)
<i>Questionnaire NOS</i>	2 (14.3)	—	3 (12.5)
<i>Chronic Liver Disease Questionnaire</i>	1 (7.1)*	1 (11.1)*	2 (8.3)
<i>EQ-5D</i>	—	2 (22.2)	2 (8.3)
PSC-PRO	—	1 (11.1)*	1 (4.2)
<i>Visual Analog Scale NOS</i>	—	1 (11.1)	1 (4.2)

Note: Bolded instruments denote instruments that were either developed in the target population or subsequently underwent psychometric testing in the target population.

*Instrument was developed with input from the target population.

†Instrument was not developed with input from the target population but later underwent psychometric testing in the target population.

and sound psychometric properties.⁽¹⁸⁾ Because the PBC-40 also includes fatigue and itch subscales with proven content validity, this instrument is an ideal patient-centered choice for studies of PBC HRQOL, fatigue, and itch. Importantly, the instrument has also been adapted into briefer versions (PBC-27, PBC-10).^(23,26) In addition to the PBC-40, the Itch Diary was recently developed based on qualitative interviews with patients with PBC, and thus is another viable option to measure PBC-induced itch.⁽¹⁹⁾ Finally, the psychometric properties of the Fatigue Impact Scale

and the Fisk Fatigue Severity Scale were evaluated for use in PBC in 2000, with the Fatigue Impact Scale being used frequently thereafter; however, because both were originally developed in other patient populations, neither can claim to be truly representative of PBC-related fatigue until further qualitative studies are conducted.^(20,21)

In contrast with PBC, far fewer studies of PRO concepts and instruments have been published for PSC. This may, in part, be due to the fact that no validated PRO instruments were developed specifically for PSC until 2018. The PSC-PRO instrument has followed guidelines for PRO measure development; more specifically, cognitive interviews were conducted with patients with PSC to validate item content.⁽²⁵⁾ However, patients from advocacy groups have stated that the PSC-PRO does not provide an accurate representation of how patients with PSC function and feel, as patients did not partake in the initial stages of development of the tool. The qualitative findings, which would provide valuable information for all PSC stakeholders, have not been published. This tool requires future testing in larger diverse samples of patients with PSC, to determine whether the overall measure is sensitive to change from disease progression or treatment. The CLDQ is a disease-specific measure of HRQOL developed for patients with various types of chronic liver disease; however, only 30% of the study cohort had PSC or PBC.⁽²²⁾ Finally, there is a new PSC HRQOL instrument currently under development that includes substantive patient engagement in every stage of development of the tool and will undergo international validation.⁽²⁷⁾

Aside from the new PSC-PRO that contains one itch item and two fatigue items, we identified no optimal choice for fatigue and pruritus measurements in PSC. Future qualitative studies should investigate patient preferences regarding pre-existing instruments to evaluate PSC-associated fatigue and itch, including the PBC-40, Fatigue Impact Scale, Fisk Fatigue Severity Scale, 5-D Itch Scale, Itch Diary, and ItchRO (developed qualitatively in children with Alagilles). The 5-D Itch measure was developed in patients with itch, but not specifically PSC, and encompasses five dimensions: duration, degree, direction, disability, and distribution. Finally, studies need to better distinguish between symptoms attributable to inflammatory bowel disease, a comorbidity that occurs in over 70% of patients with PSC, and those attributed to PSC.

This study revealed significant gaps in the extant literature and highlights implications for future research. For instance, most PRO studies used visual analog and numeric rating scales, yet we identified no published scientific justification for these scales based on patient preferences. Moreover, there was significant heterogeneity in the types of visual and numeric rating scales, making it difficult to compare results in this literature. Therefore, a reduced reliance on these types of instruments in future studies is recommended until more empirical support is demonstrated. More studies in special populations (e.g., pediatric, pregnant, liver transplantation) are needed to understand patients' perspectives, needs, and priorities to inform study outcomes and measures. There may be opportunities to modify and validate the National Institutes of Health Patient-Reported Outcome Measurement System measures in PBC and PSC, as these measures underwent robust qualitative development and psychometric testing in diverse populations.

Most clinical trials in PBC or PSC did not measure AEs from the patients' point of view and did not use PRO instruments. As a result, important information regarding the harms of treatments may be undetected and underrecognized. We acknowledge that these results likely reflect that patient-centered reporting of AEs is not a part of the current culture of conducting clinical trials. However, it is precisely for this reason that we believe existing practices should adopt a more patient-centered approach. Evaluating AEs systematically using PRO instruments is another way to enhance patient-focused drug development. Recently, the National Cancer Institute spearheaded the development of a PRO version of the Common Terminology Criteria for Adverse Events, called the "PRO-CTCAE," as a more accurate method of detecting AEs during oncology drug trials.⁽²⁸⁾ The items were identified and developed using cognitive interviews with patients receiving active cancer treatment; furthermore, they underwent validation and extensive psychometric testing.^(29,30) The PRO-CTCAE contains 78 patient-reported AEs rated on presence, frequency, severity, and life interference. A recent study highlighted significant discordance between clinician-graded AEs and the PRO-CTCAE, demonstrating the complementary and added value of the patient instrument.⁽³¹⁾ A high-yield contribution to the field could be to imbed qualitative studies into PBC or PSC drug trials to

develop a modified version of the PRO-CTCAE for these patients. As a way to facilitate patient-centered reporting of AEs, electronic modalities (e.g., tablets, cellphones, websites) could be used for reporting of AEs; participants could use validated instruments to report symptoms or experiences they had during the clinical trial.

It is imperative that more qualitative or mixed methods studies be conducted to move the field of PROs in PBC and PSC forward. We identified only three qualitative and three mixed-methods articles that described the experiences of patients with PBC^(12,32-36); astonishingly, no qualitative studies have been published describing the experiences of patients with PSC. We acknowledge that there may be a bias against publishing qualitative studies, and researchers may not be incentivized to conduct these types of works. However, qualitative studies are absolutely essential to deepen our understanding of patient experiences, perspectives, needs, and priorities; this is especially true in the setting of studying different medical treatments.⁽³⁷⁻³⁹⁾ Toward this end, clinical investigators are encouraged to familiarize themselves with the systematic, step-by-step approaches for conducting formative, qualitative work needed to develop or modify accompanying PRO instruments. In 2019, the FDA published a series of four guidance documents to facilitate patient-focused drug development.⁽⁴⁰⁾ The new guidance includes best practices for qualitative research to gain insight into patient experiences and needs. Additionally, the International Society for Pharmacoeconomics and Outcomes has also published two best practice papers for developing new, or modifying pre-existing, PRO instruments.^(16,41) Collectively these guidances provide systematic approaches to conduct qualitative work that can move the field forward.

In conclusion, this systematic review identified the most commonly reported PRO concepts and PRO instruments published in the PBC and PSC literature. PRO instruments were used to measure these concepts only half the time, and most instruments were not validated in the target populations. Given significant gaps in our current knowledge regarding patient experiences and preferences, more qualitative studies and systematic approaches for incorporating the patient voice are needed in current research endeavors, including clinical trials, among patients living with PBC and PSC.

REFERENCES

- Selmi C, Gershwin ME, Lindor KD, Worman HJ, Gold EB, Watnik M, et al. Quality of life and everyday activities in patients with primary biliary cirrhosis. *Hepatology* 2007;46:1836-1843.
- Rannard A, Buck D, Jones DEJ, James OFW, Jacoby A. Assessing quality of life in primary biliary cirrhosis. *Clin Gastroenterol Hepatol* 2004;2:164-174.
- Untas A, Boujut E, Corpechot C, Zenasni F, Chazouillères O, Jaury P, et al. Quality of life and illness perception in primary biliary cirrhosis: a controlled cross-sectional study. *Clin Res Hepatol Gastroenterol* 2015;39:52-58.
- Younossi ZM, Kiwi ML, Boparai N, Price LL, Guyatt G. Cholestatic liver diseases and health-related quality of life. *Am J Gastroenterol* 2000;95:497-502.
- Mells GF, Pells G, Newton JL, Bathgate AJ, Burroughs AK, Heneghan MA, et al. Impact of primary biliary cirrhosis on perceived quality of life: the UK-PBC national study. *Hepatology* 2013;58:273-283.
- Newton JL, Elliott C, Frith J, Ghazala C, Pairman J, Jones DEJ. Functional capacity is significantly impaired in primary biliary cirrhosis and is related to orthostatic symptoms. *Eur J Gastroenterol Hepatol* 2011;23:566-572.
- Gotthardt DN, Rupp C, Bruhin M, Schellberg D, Weiss KH, Stefan R, et al. Pruritus is associated with severely impaired quality of life in patients with primary sclerosing cholangitis. *Eur J Gastroenterol Hepatol* 2014;26:1374-1379.
- Hegade VS, Mells GF, Fisher H, Kendrick S, DiBello J, Gilchrist K, et al. Pruritus is common and undertreated in patients with primary biliary cholangitis in the United Kingdom. *Clin Gastroenterol Hepatol* 2019;17:1379-1387.e3.
- Hegade VS, Bolier R, Oude Elferink RP, Beuers U, Kendrick S, Jones DE. A systematic approach to the management of cholestatic pruritus in primary biliary cirrhosis. *Frontline Gastroenterol* 2016;7:158-166.
- Newton JL, Gibson GJ, Tomlinson M, Wilton K, Jones D. Fatigue in primary biliary cirrhosis is associated with excessive daytime somnolence. *Hepatology* 2006;44:91-98.
- Newton JL, Pairman J, Sutcliffe K, Wilton K, Jones DEJ. A predictive model for fatigue and its etiologic associations in primary biliary cirrhosis. *Clin Gastroenterol Hepatol* 2008;6:228-233.
- Blackburn P, Freeston M, Baker CR, Jones DEJ, Newton JL. The role of psychological factors in the fatigue of primary biliary cirrhosis. *Liver Int* 2007;27:654-661.
- Stanca CM, Bach N, Krause C, Tandon N, Freni MA, Gutierrez JA, et al. Evaluation of fatigue in U.S. patients with primary biliary cirrhosis. *Am J Gastroenterol* 2005;100:1104-1109.
- Al-Harthi N, Kumagi T, Coltescu C, Hirschfield GM. The specificity of fatigue in primary biliary cirrhosis: evaluation of a large clinic practice. *Hepatology* 2010;52:562-570.
- U.S. Food and Drug Administration (FDA). Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. 2009. <https://www-fda-gov.libproxy.lib.unc.edu/media/77832/download>. Accessed August 6, 2019.
- Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report. Part 1: Eliciting concepts for a new PRO instru. *Value Heal* 2011;14:967-977.
- U.S. Food and Drug Administration (FDA). Methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments discussion document for

- patient-focused drug development. 2018. <https://www.fda.gov/media/116276/download>. Accessed March 20, 2020.
- 18) **Jacoby A, Rannard A**, Buck D, Bhala N, Newton JL, James OF, 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;54:1622-1629.
 - 19) Martin ML, Stassek L, Blum SI, Joshi AV, Jones D. Development and adaptation of patient-reported outcome measures for patients who experience itch associated with primary biliary cholangitis. *J Patient-Reported Outcomes* 2019;3:2.
 - 20) Huet PM, Deslauriers J, Tran A, Faucher C, Charbonneau J. Impact of fatigue on the quality of life of patients with primary biliary cirrhosis. *Am J Gastroenterol* 2000;95:760-767.
 - 21) Prince MI, James OF, Holland NP, Jones DE. Validation of a fatigue impact score in primary biliary cirrhosis: towards a standard for clinical and trial use. *J Hepatol* 2000;32:368-373.
 - 22) Younossi ZM, Guyatt G, Kiwi M, Boparai N, King D. Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease. *Gut* 1999;45:295-300.
 - 23) **Montali L, Tanaka A**, Riva P, Takahashi H, Cocchi C, Ueno Y, et al. A short version of a HRQoL questionnaire for Italian and Japanese patients with primary biliary cirrhosis. *Dig Liver Dis* 2010;42:718-723.
 - 24) Raszeja-Wyszomirska J, Wunsch E, Krawczyk M, Rigopoulou EI, Bogdanos D, Milkiewicz P. Prospective evaluation of PBC-specific health-related quality of life questionnaires in patients with primary sclerosing cholangitis. *Liver Int* 2015;35:1764-1771.
 - 25) Younossi ZM, Afendy A, Stepanova M, Racila A, Nader F, Gomel R, et al. Development and validation of a primary sclerosing cholangitis-specific patient-reported outcomes instrument: the PSC PRO. *Hepatology* 2018;68:155-165.
 - 26) Alrubaiy L, Mellis G, Flack S, Bosomworth H, Hutchings H, Williams J, et al. PBC-10: a short quality of life measure for clinical screening in primary biliary cholangitis. *Aliment Pharmacol Ther* 2019;50:1223-1231.
 - 27) PSC Support. UK-PSC quality of life update. 2019. <https://www.pscsupport.org.uk/uk-psc-quality-of-life-update/>. Accessed March 20, 2020.
 - 28) Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014;106:dju244.
 - 29) Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, et al. Cognitive interviewing of the US national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *Qual Life Res* 2014;23:257-269.
 - 30) Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and reliability of the U.S. national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *JAMA Oncol* 2015;1:1051-1059.
 - 31) Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. *Support Care Cancer* 2016;24:3669-3676.
 - 32) Lasker JN, Sogolow ED, Sharim RR. The role of an online community for people with a rare disease: content analysis of messages posted on a primary biliary cirrhosis mailing list. *J Med Internet Res* 2005;7:e10.
 - 33) Jorgensen R. A phenomenological study of fatigue in patients with primary biliary cirrhosis. *J Adv Nurs* 2006;55:689-697.
 - 34) Montali L, Frigerio A, Riva P, Invernizzi P. "It's as if PBC didn't exist": the illness experience of women affected by primary biliary cirrhosis. *Psychol Health* 2011;26:1429-1445.
 - 35) Sogolow ED, Lasker JN, Sharim RR, Weinrieb RM, Sass DA. Stigma and liver disease. *Illn Cris Loss* 2010;18:229-255.
 - 36) Robinson L, Newton JL, Jones D, Dawson P. Promoting self-management and adherence with strength and balance training for older people with long-term conditions: a mixed-methods study. *J Eval Clin Pr* 2014;20:318-326.
 - 37) Evon DM, Golin CE, Stoica T, Jones RE, Willis SJ, Galanko J, et al. What's important to the patient? Informational needs of patients making decisions about hepatitis C treatment. *Patient* 2017;10:335-344.
 - 38) Evon DM, Golin CE, Bonner JE, Grodensky C, Velloza J. Adherence during antiviral treatment regimens for chronic hepatitis C: a qualitative study of patient-reported facilitators and barriers. *J Clin Gastroenterol* 2015;49:e41-e50.
 - 39) U.S. Food and Drug Administration (FDA). Patient-focused drug development glossary. 2018. <https://www.fda.gov/drugs/development-approval-process-drugs/patient-focused-drug-development-glossary>. Accessed March 16, 2020.
 - 40) U.S. Food and Drug Administration (FDA). FDA patient-focused drug development guidance series for enhancing the incorporation of the patient's voice in medical product development and regulatory decision making. 2019. <https://www.fda.gov/drugs/developmentapprovalprocess/ucm610279.htm>. Accessed March 20, 2020.
 - 41) Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report. Part 2: Assessing respondent understanding. *Value Heal* 2011;14:978-988.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1567/supinfo.