DOI: 10.1111/1475-6773.13187

RESEARCH ARTICLE

HSR Health Services Research

A systematic review of the validity and reliability of patient-reported experience measures

Claudia Bull BNutr[1,2] | Joshua Byrnes B Comm, B Econ, M Econ Stud, M Hlth Econ, PhD[1,2] | Ruvini Hettiarachchi BDS, MSc, MD[1,2] | Martin Downes MVB, PhD, PGCertPH[1,2]

[1]Centre for Applied Health Economics (CAHE), Griffith University, Brisbane, Queensland, Australia

[2]Menzies Health Institute Queensland (MHIQ), Brisbane, Queensland, Australia

Correspondence
Claudia Bull, BNutr, Centre for Applied Health Economics (CAHE), Griffith University, 170 Kessels Road, Nathan, QLD 4111, Australia.
Email: claudia.bull@griffithuni.edu.au

Abstract

Objectives: To identify patient-reported experience measures (PREMs), assess their validity and reliability, and assess any bias in the study design of PREM validity and reliability testing.

Data Sources/Study Setting: Articles reporting on PREM development and testing sourced from MEDLINE, CINAHL and Scopus databases up to March 13, 2018.

Study Design: Systematic review.

Data Collection/Extraction Methods: Critical appraisal of PREM study design was undertaken using the Appraisal tool for Cross-Sectional Studies (AXIS). Critical appraisal of PREM validity and reliability was undertaken using a revised version of the COSMIN checklist.

Principal Findings: Eighty-eight PREMs were identified, spanning across four main health care contexts. PREM validity and reliability was supported by appropriate study designs. Internal consistency (n = 58, 65.2 percent), structural validity (n = 49, 55.1 percent), and content validity (n = 34, 38.2 percent) were the most frequently reported validity and reliability tests.

Conclusions: Careful consideration should be given when selecting PREMs, particularly as seven of the 10 validity and reliability criteria were not undertaken in ≥50 percent of the PREMs. Testing PREM responsiveness should be prioritized for the application of PREMs where the end user is measuring change over time. Assessing measurement error/agreement of PREMs is important to understand the clinical relevancy of PREM scores used in a health care evaluation capacity.

KEYWORDS
health care organization and systems, reliability, survey research and questionnaire design, systematic reviews/meta-analyses, validity

1 | INTRODUCTION

Patient-reported experience measures (PREMs) are tools that capture "what" happened during an episode of care, and "how" it happened from the perspective of the patient.[1-3] PREMs differ from patient-reported outcome measures (PROMs), which aim to measure patients' health status,[4] and the more subjective patient satisfaction measures, which are an indication of how well a patient's expectations were met,[5] a benchmark which is criticized for being too heavily influenced by past health care encounters.[6]

Patient-reported experience measures are gaining attention as an indicator of health care quality and can provide information

regarding the patient-centeredness of existing services as well as areas for potential improvement regarding health care delivery.[7] The purpose of employing PREMs is consistent with the Institute of Medicine's (IOM) definition of health care quality, defined as care that is patient-centered, effective, efficient, timely, and equitable.[8] In recent years, PREMs have been used to inform pay-for-performance (P4P) and benchmarking schemes, adjunct with other health care quality domains, including clinical quality/effectiveness, health information technology, and resource use.[9,10] Such schemes see health care services financially rewarded for their performance across these domains of health care quality, as opposed to the traditional fee-for-service payment system, which may inadvertently promote low-value care.[10,11]

While there is evident merit behind utilizing PREMs in health care quality evaluations, there remains some conjecture regarding their use. Manary and colleagues[12] identify three main limitations expressed by critics of PREMs. Firstly, patient-reported experience is largely seen as congruent with terms such as "patient satisfaction" and "patient expectation," both of which are subjective terms that can be reflective of judgments on the *adequacy* of health care and not the *quality*.[12-14] Secondly, PREMs may be confounded by factors not directly related to the quality of health care experienced by the patient, such as health outcomes.[12] And finally, PREMs can be a reflection of patients' preconceived health care "ideals" or expectations and not their actual care experience.[12] All three limitations are indicative of a blurring of concept boundaries and inappropriate interchanging of concepts. While this is not unique to PREMs, it does suggest a low level of concept maturity regarding patient-reported experiences[15] and, consequently, is an area of research that warrants greater attention.

Despite these limitations, PREMs have gained international recognition as an indicator of health care quality. This is largely because: (a) they enable patients to comprehensively reflect on the interpersonal aspects of their care experience[16]; (b) they can be utilized as a common measure for public reporting, benchmarking of institutions/centers and health care plans[10]; and (c) they can provide patient-level information that is useful in driving service quality improvement strategies.[17,18]

Understanding the validity and reliability of PREMs is integral to the appropriate selection of instruments for quality assessment of health care services, in conjunction with other aspects, such as the clinical relevance of an instrument and the domains of patient-reported experience that the PREM covers. Validity refers to the ability of an instrument to measure what it intends to measure, and reliability refers to the ability of an instrument to produce consistent results under similar circumstances, as well as to discriminate between the performance of different providers.[19,20] It is important to assess these properties in order to understand the risk of bias that may arise in employing certain instruments[21] and whether instruments are suitable for capturing patient-reported experience data.

While two previously published systematic reviews have examined the psychometric testing of PREMs, one related to PREMs for inpatients,[16] and the other for emergency care service provision,[22] there has been no comprehensive examination of the tools available across a range of health care contexts. The aim of this systematic review was to identify PREMs, assess their validity and reliability, and assess any bias in the study design of PREM validity and reliability testing, irrespective of the health care context the PREMs are designed to be used in.

1.1 | Objectives

1. To identify existing tools for measuring patient-reported experiences in health care, irrespective of the context
2. To critically appraise bias in the study design employed in PREM validity and reliability testing, and
3. To critically appraise the results of validity and reliability testing undertaken for these PREMs.

2 | METHODS

In conducting this systematic review, the authors conformed to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.[23] This review was registered with PROSPERO (registration number: CRD42018089935).

2.1 | Search strategy and eligibility criteria

The databases searched were MEDLINE (Ovid), CINAHL Plus with full text (EBSCOhost), and Scopus (Elsevier). No date restriction was applied to the search strategy; records were searched up to March 13, 2018. Patient "satisfaction" was included in search terms in order to not to limit our results as there is a blurring of these terms and some PREMs may be labeled as satisfaction measures.

Articles were included in the systematic review if they met all the following criteria:

- Described the development and evaluation of PREMs
- Published in English
- Full-text published in peer-reviewed journals
- Labeled as a satisfaction questionnaire, but framed around measuring patients' experiences (eg, the Surgical In-Patient Satisfaction (SIPs) instrument[24])

Articles were excluded if they met any of the following criteria:

- Instruments labeled as a satisfaction questionnaire which were: (a) framed around measuring patient levels of satisfaction; (b) inclusive of a global satisfaction question or visual analogue scale; and (c) developed based on satisfaction frameworks or content analyses
- Patient expectation questionnaires

- Quality of care questionnaires
- Patient participation questionnaires
- Related to patient experiences of a specific treatment or intervention (eg, insulin devices, hearing aids, food services, anesthesia, and medication/pharmaceutical dispensary) or specific program (eg, education programs)
- Measuring emotional care received by patients (eg, empathy)
- Studies where PREMs were completed entirely by proxy (completed by populations *not* receiving the care); however, if proxy-reported data comprised only a small proportion of data collected (patient-reported data also reported), then the study was still included
- Quality improvement initiatives
- Patient attitude scales
- Checklists
- Patient experience questionnaires comprised of a single domain, or
- PREMs superseded by a more up-to-date version of the same PREM with corresponding updated validity and reliability testing

The full search strategy for each database is provided in Appendix S1. All references were imported into EndNote (Version 8, Clarivate Analytics), and duplicates were removed. Two reviewers independently screened paper titles and abstracts for inclusion. Where the title and abstract were not informative enough to make a decision, the full-text article was reviewed. Figure 1 depicts the PRISMA flow diagram of this process. The two reviewers handled disagreements regarding article inclusion or exclusion. Where a decision could not be made, a third reviewer adjudicated the final decision. Reference list handsearching was also employed for the identification of PREMs not identified through database searching, and updates for PREMs originally identified through the database searching.

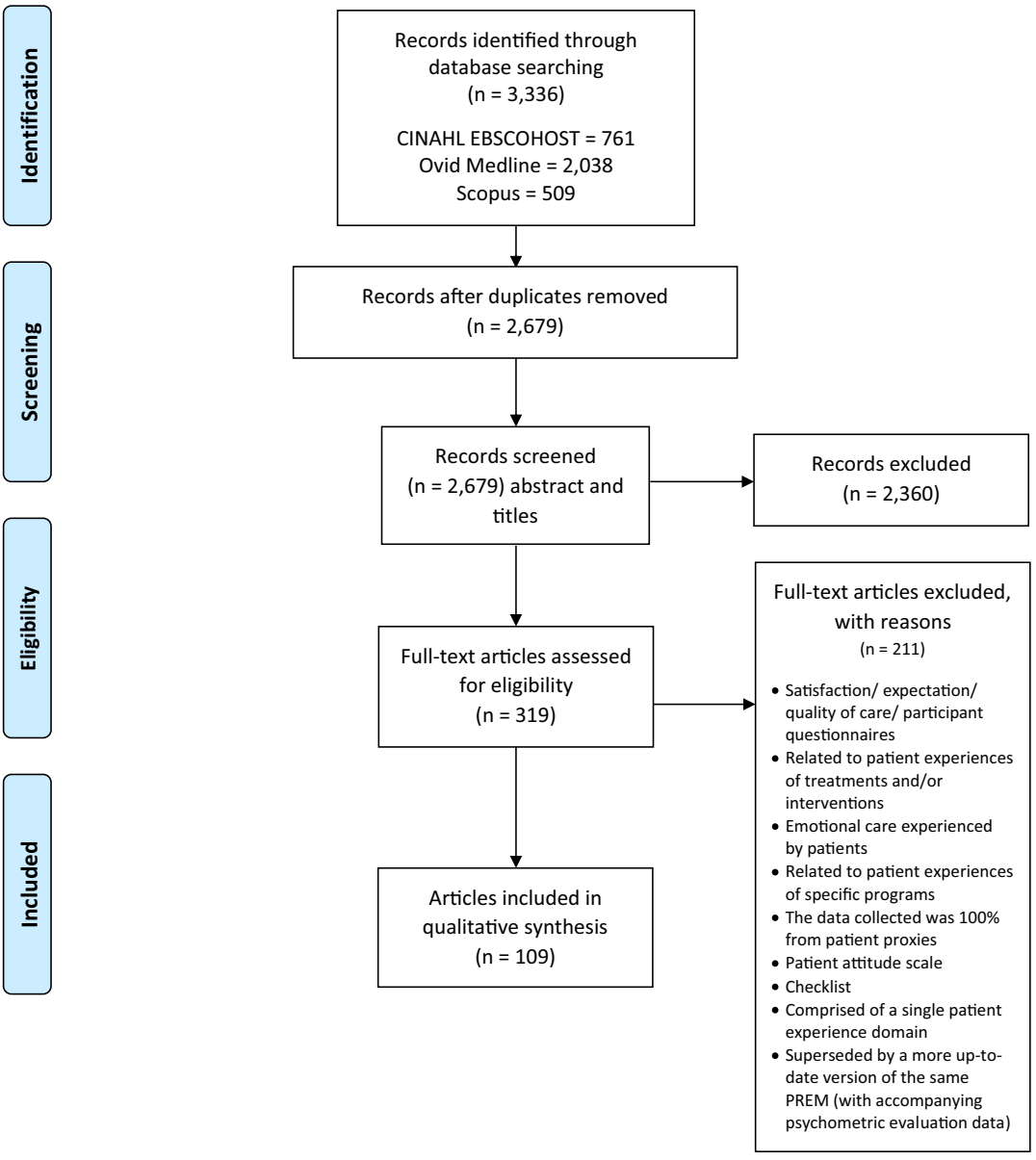

FIGURE 1 PRISMA diagram of patient-reported experience measure search [Color figure can be viewed at wileyonlinelibrary.com] [Color figure can be viewed at wileyonlinelibrary.com]

2.2 | Data extraction

Descriptive data were independently extracted from the included articles by two reviewers into a standardized excel extraction form (refer to Appendix S2). Discrepancies in the extracted data were discussed between the two reviewers, or adjudicated by a third if necessary. If there was insufficient information in the full-text article regarding the validity and reliability testing undertaken, the article was excluded.

2.3 | Critical appraisal

To critically appraise bias in the study design employed in PREM validity and reliability testing, the Appraisal tool for Cross-Sectional Studies (AXIS)[25] was used. This is a 20-item appraisal tool developed in response to the increase in cross-sectional studies informing evidence-based medicine and the consequent importance of ensuring that these studies are of high quality and low bias.[25] The purpose of employing the AXIS tool in the present systematic review was to ensure that the results of PREM validity and reliability testing were supported by appropriate study designs and thus able to be interpreted as a robust representation of how valid and/or reliable a PREM is. The AXIS assesses the quality of cross-sectional studies based on the following criteria: clarity of aims/objectives and target population; appropriate study design and sampling framework; justification for the sample size; measures taken to address nonresponders and the potential for response bias; risk factors/outcome variables measured in the study; clarity of methods and statistical approach; appropriate result presentation, including internal consistency; justified discussion points and conclusion; discussion of limitations; and identification of ethical approval and any conflicts of interest.[25]

The scoring system conforms to a "yes," "no," or "do not know/comment" design. PREMs were categorized into quartiles: >15 AXIS criteria met, 10-15 AXIS criteria met, 5-9 AXIS criteria met, and ≤4 AXIS criteria met. The AXIS tool was used to appraise the most recent publication for each PREM as this was also reflective of the most recent version of validity and reliability testing that the PREM that had undergone.

To assess the validity and reliability testing undertaken for PREMs included in this review, we employed a revised version of the COSMIN checklist (COnsensus-based Standards for the selection of health status Measurement INstruments) published in a recent systematic review of quality in shared decision making (SDM) tools.[19] These criteria is comprised of 10 psychometric measurement properties and subproperties, including internal consistency; reliability; measurement error/agreement; validity (inclusive of content validity, construct validity [structural validity, hypothesis testing, and cross-cultural validity]; and criterion validity); responsiveness; and item response theory (IRT). Appendix S3 provides definitions for each of these measurement properties and identifies the appraisal parameters used to assess them.[26]

Reporting of these measurement properties conforms to the following: "+" (meets criteria), "–" (does not meet criteria), or "?" (unclear or missing information). These scores were numerically coded, and PREMs were ranked within their corresponding context(s) (refer to Appendix S4). Where more than one article was identified for the validity and reliability testing of a PREM, all articles were used to critically appraise the PREM. If the same criteria were assessed in separate studies for a given PREM and provided conflicting results (eg, a "+" and a "–" score), then the more favorable result was recorded.

Appraisals with both tools were undertaken by one author. A sample of the revised COSMIN checklist appraisal data was cross-checked with a second reviewer. A Kappa measure was used to assess the level of inter-rater agreement. A Kappa value of 0.5 depicted moderate agreement, >0.7 good agreement, and >0.8 very good agreement.[27]

3 | RESULTS

A total of 88 PREMs were identified through the systematic literature search. Greater than one-third of these instruments were contextually designed for inpatient care services (36.4 percent), 23.9 percent for primary care services and 12.5 percent for outpatient care services. Table 1 depicts the other contexts and conditions covered by the PREMs. Roughly 20 percent of instruments were developed in the UK, while other countries included the United States (19.3 percent), Norway (14.8 percent), and the Netherlands (13.6 percent). The most common mode of PREM administration was postal (45.7 percent), followed by face to face (33.1 percent), telephone (13.6 percent), and electronic (7.6 percent). The earliest PREMs detected through the systematic search were developed in 1993.[28,29] The median number of items per PREM was 27 (IQR: 21-35; range: 4-82), and the median number of domains was 5 (IQR: 4-7; distribution: 2-13). Extracted data can be identified in Appendix S2.

A proxy, not the recipient of care, completed PREMs on the behalf of patients in 11.4 percent of the PREMs. This was typically only for a small portion (10-12 percent) of any given study's population. Over 40 percent of the PREMs were developed and tested in languages other than English. Few papers discuss formal translation processes being undertaken for PREMs.

3.1 | AXIS critical appraisal

Table 2 identifies that 63 (70.5 percent) of the papers reporting on PREMs met >15 AXIS criteria. Over a quarter of studies met 10-15 criteria (28.4 percent), and 1.1 percent (n = 1) met 5-9 criteria. No PREM met ≤4 AXIS criteria. The median number of "yes" scores was 16 (IQR: 15-17). The lowest scoring of all PREMs answered "yes" to only five of the 20 AXIS questions.[30] The highest scoring PREM answered "yes" to all questions.[31] Appendix S5 presents the AXIS results for all PREMs from highest to lowest number of AXIS criteria met.

TABLE 1 PREMs identified by individual, condition, setting, and country-specific context

Context[a]	Number of PREMs (%)
Individual-specific	
Child/adolescent care[57,58]	2 (2.3)
Low-income[59]	2 (2.3)
Homeless[60]	1 (1.1)
Not individual-specific	83 (94.3)
Condition-specific	
Mental health[61-73]	10 (11.4)
Palliative care/cancer[36,74-76]	5 (5.7)
Renal (including dialysis)[77-79]	3 (3.4
Rheumatoid arthritis[80-82]	3 (3.4)
Substance dependence[64,65,83]	2 (2.3)
Trauma[84-87]	2 (2.3)
Chronic disease[88]	1 (1.1)
Cystic fibrosis[89]	1 (1.1)
Maternity[90,91]	1 (1.1)
Parkinson's disease[92]	1 (1.1)
Not condition-specific	59 (67.1)
Setting	
Inpatient services[28,30,31,59,61,63,66,72,73,76,81,93-108]	32 (36.4)
Day surgery[28]	1 (1.1)
Rehabilitation[81,99]	2 (2.3)
Preoperative[109]	1 (1.1)
Postoperative[47,110]	3 (3.4)
Primary care services[34,46,60,69,88,111-132]	21 (23.9)
Medical home[133]	1 (1.1)
Out-of-hours care[113,118]	2 (2.3)
Home care[29]	1 (1.1)
Outpatient services[59,67,68,71,75,76,109,134-138]	11 (12.5)
Accident and emergency department services[139-142]	3 (3.4)
Dental services[143,144]	2 (2.3)
Integrated care services[145,146]	2 (2.3)
Not specified[35,37,147-155]	5 (5.7)
Not setting-specific	1 (1.1)
Country	
UK[28,58,66,69,75,80,87,98,99,104,112-114,119,122,124,126,129,139,142,147,148]	18 (20.5)
USA[29,34,35,57,60,64,65,74,78,79,89,116,117,123,127,131,133,143,146,150-165]	17 (19.4)
Norway[31,37,61,67,71,76,81,83,102,103,118,121,130,135]	13 (14.8)
Netherlands[36,62,77,82,90-93,105,140,141,166,167]	13 (14.8)
Australia[70,115,120,144]	4 (4.5)
Spain[88,94,96,109]	4 (4.5)
Canada[47,84-86,137,138]	3 (3.4)
Hong Kong[24,95,106-108]	3 (3.4)
Sweden[68,72,73,122]	3 (3.4)
China[30,136]	2 (2.3)
Ethiopia[59]	2 (2.3)
Germany[63,145]	2 (2.3)
Europe[b,46,111,125,132]	1 (1.1)
France[100]	1 (1.1)
Saudi Arabia/UAE[149]	1 (1.1)
Taiwan[134]	1 (1.1)
Total	**88 (100)**

Abbreviations: PREM, patient-reported experience measure; UAE, United Arab Emirates; UK, United Kingdom; USA, United States of America.

[a]Some tools were embedded across contexts.

[b]One study was conducted across 14 countries in Europe.

Appendix S6 identifies that all studies we assessed as presenting clear study aims and utilizing appropriate study designs to answer their research questions (Q1 and Q2). Greater than 95 percent of PREMs appropriately sampled participants to be representative of the target population under investigation (Q5). Over 95 percent of the studies reported that there was no conflict of interest related to a funding source and the interpretation of results (Q19). Questions 13 (potential for response rate bias), 14 (description of nonresponders), and 20 (attainment of ethical approval or participant consent) were the criteria least frequently met by PREM papers.

3.2 | Revised COSMIN checklist validity and reliability appraisal

Appendix S4 details the validity and reliability testing undertaken for the PREMs according to the revised COMSIN checklist. PREMs are ranked within their specified contexts according to the number of positive results obtained for the validity and reliability tests. Inter-rater reliability between two assessors for a portion of the COSMIN checklist appraisals was κ = 0.761, indicative of good agreement.

Some validity and reliability tests were undertaken more often than others (Table 3). The three psychometric tests most commonly meeting "+" criteria were internal consistency (n = 58, 65.9 percent), structural validity (n = 49, 55.7 percent), and content validity (n = 33, 37.5 percent). Seven of the 10 revised COSMIN checklist criteria were not undertaken in ≥50 percent of the PREMs: (a) reliability (n = 44, 50.0 percent); (b) hypotheses testing (n = 53, 60.2 percent); (c) cross-cultural validity (n = 65, 73.9 percent); (d) criterion validity (n = 79, 89.8 percent); (e) responsiveness (82, 93.2 percent); (f) item

TABLE 2 PREMs categorized according to proportion of AXIS criteria met

Total AXIS score	PREM	n (%)
>15 AXIS criteria met	NORPEQ[31]; AEDQ[139]; CACHE[94]; CAHPS HIT[123]; CQI-CSD[105]; CQI-Hip Knee[110]; MPOC-A[47]; OPEQ[135]; OPEQ-China[136]; PEPAP-Q[109]; POPEQ[71]; QTAC-PREM[86]; SF-HKIEQ[108]; AIPS[104]; CAHPS Dental plan[143]; CQI A&E[141]; CSS[29]; EUROPEP[125]; González, Quintana, Bilbao, Escobar, Aizpuru, Thompson, Esteban, San Sebastián, De la Sierra[96]; OPQ[113]; PCQ-PD[92]; PEACS 1.0[145]; PEQ-GP[121]; PEQ-ITSD[83]; PEQ-OHC[118]; PFQ[134]; PIPEQ-OS[61]; QPC[73]; Re-PEQ[81]; VOICE[66]; ADAPT[57]; Bruyneel, Van Houdt, Coeckelberghs, Sermeus, Tambuyzer, Cosemans, Peeters, Van den Broeck, Weeghmans, Vanhaecht[62]; CABHS[65]; CAHPS CC[155]; CAHPS PCMH[133]; CPEQ[a,76]; CPEQ[b,76]; CQI-Cataract[167]; CQI-RA[82]; GPAQ[124]; GPPS[112]; I-PAHC[c,59];LifeCourse[74]; MCQ[75]; O-PAHC[d,59]; PEPAC[166]; PEQ MH[69]; ReproQ[91]; Walker, Stewart, Grumbach[146]; Bruyneel, Tambuyzer, Coeckelberghs, De Wachter, Sermeus, De Ridder, Ramaekers, Weeghmans, Vanhaecht[93]; CAHPS Health Plan[151]; CEO-MHS[70]; ChASE[58]; Homa, Sabadosa, Nelson, Rogers, Marshall[89]; ICEQ[87]; IDES[63]; IEXPAC[88]; Labarère, Fourny, Jean-Phillippe, Marin-Pache, Patrice[100]; NREQ[99]; PSQ MD[138]; Steine, Finset, Laerum[130]; UCSQ[142]	62 (70.5)
10-15 criteria met	ACES[127]; Black, Sanderson[28]; CAHPS C&G[131]; CQI-CHD[77]; CQI-PHHD[77]; GS-PEQ[37]; HKIEQ[107]; HQ[68]; I-PEQ CHD[98]; IPQ[119]; PAIS[120]; PCQ-H[60]; PEQ[103]; PESS[115]; SIPS[24]; CAHPS ICH[78]; DPQ[144]; Drain[116]; HCAHPS[157]; howRwe[148]; Malott, Fulton, Rigamonti, Myers[149]; PREM RA and Other[80]; Picker MSD[122]; PPE-15[97]; CQI-PC[36]	25 (28.4)
5-9 criteria met	PEES-50[30]	1 (1.1)
≤4 criteria met	Nil	0 (0)

Abbreviations: AXIS, Appraisal tool for Cross-Sectional Studies; PREM, patient-reported experience measure.
[a]Outpatient version of CPEQ.
[b]Inpatient version of CPEQ.
[c]Reports I-PAHC.
[d]Reports O-PAHC.

response theory (n = 84, 95.5 percent); and (g) measurement error/agreement (n = 86, 97.8 percent). None of the studies undertook testing for all 10 validity and reliability criteria.

4 | DISCUSSION

The purpose of this systematic review was threefold: to identify and describe peer-reviewed PREMs, irrespective of their contextual basis; to critically appraise PREM validity and reliability; and to critically appraise any bias in the study design of PREM validity and reliability testing. It is integral to understand whether PREMs have been subject to rigorous validity and reliability testing as this reflects whether an instrument is able to appropriately capture patient-reported experiences of health care. In turn, it is also important to ensure that the results of PREM validity and reliability testing are underpinned by a rigorous study design so that readers can be assured that the validity and reliability results are a robust representation of how valid and/or reliable that PREM is. To our knowledge, this is the first systematic review to examine PREMs across a range of health care contexts and settings.

This systematic review identified a total of 88 PREMs. Interestingly, roughly 20 percent of the identified PREMs were developed from 2015 onwards, and a quarter of all PREMs received some form of additional validity and reliability testing in this time frame as well. Given that 1993 was the earliest PREM development year identified through the search strategy, this indicates a significant increase in the desire for instruments that measure patient experiences.

Generally, the PREMs identified in this systematic review reflect a heavy emphasis on measuring singular events of health care. The Institute of Medicine's (IOM) 2001 report on crossing the quality chasm identified that despite a significant increase in chronic and complex conditions, health care systems are still devoted to acute episodes of care.[32] Overwhelmingly, this sentiment still reigns true today, despite efforts to promote greater integration and coordination of care across and within health care services, as well as patient-centric, high-quality health care.[8,33] For example, this systematic review identified only one peer-reviewed PREM targeting chronic disease holistically (as opposed to a singular disease focus) and two PREMs focusing on the integration of care. Most PREMs related to short-term care episodes, largely in the hospital setting, though there are PREMs (eg, the CG-CAHPS[34] and health plan CAHPS[35]) that examine patient experiences of care delivered over 6- to 12-month periods. By developing and utilizing PREMs that maintain a single event, unidimensional focus of health care, we are inhibiting our ability to strive for international health care goals related to reducing health care fragmentation, and optimizing continuity, coordination, and quality within and between services. Consequently, future PREM development should aim to capture patient experiences of the continuity and coordination within and between their health care services and providers in order to mirror international shifts toward greater health care integration.

Encouragingly, nearly all PREM evaluation papers met ≥10 AXIS criteria (98.9 percent). Furthermore, all papers possessed appropriate study designs for their stated aims, and >95 percent of papers demonstrated appropriate sampling of participants to be representative of the target population under investigation. One PREM,[30]

TABLE 3 Psychometric quality of PREMs according to the Revised COSMIN checklist

Psychometric quality criteria	Criteria met, n (%)	Criteria not met, n (%)	Unknown or unclear information, n (%)
Internal consistency	58 (65.9)	18 (20.5)	12 (13.6)
Reliability	18 (20.5)	26 (29.5)	44 (50.0)
Measurement error/ agreement	1 (1.1)	1 (1.1)	86 (97.8)
Content validity	33 (37.5)	0 (0)	55 (62.5)
Construct validity			
Structural validity	49 (55.7)	6 (6.8)	33 (37.5)
Hypotheses testing	21 (23.9)	14 (15.9)	53 (60.2)
Cross-cultural validity	13 (14.8)	10 (11.4)	65 (73.8)
Criterion validity	3 (3.4)	6 (6.8)	79 (89.8)
Responsiveness	4 (4.5)	2 (2.3)	82 (93.2)
IRT	3 (3.4)	1 (1.1)	84 (95.5)

Abbreviations: COSMIN, COnsensus Standards for the selection of health Measurement INstruments; IRT, item response theory; PREM, patient-reported experience measure.

however, met only five out of 20 AXIS criteria, implying that this PREM should undergo further evaluative testing prior to use in patient experience evaluations. Generally, the results of the AXIS critical appraisal indicate that the study designs underpinning PREM validity and reliability testing were sound.

Unlike the recent systematic review of hospital-related PREMs[16] where all instruments presented some form of validity and reliability testing, we identified two PREMs (CQI-PC and GS-PEQ) that did not present any testing in accordance with the revised COSMIN checklist.[36,37] This was either a consequence of not having done the testing, not presenting clear enough information to be scored "+" or "–," or not having published this information in the peer-reviewed literature. Evidently, both the CQI-PC and GS-PEQ instruments require further validity and reliability testing before being used in patient experience evaluations.

The most frequently undertaken reliability and validity criteria that also received positive results included internal consistency (a measure of reliability), structural validity, and content validity. This ultimately indicates that most PREMs measure the concept which they set out to measure, and do so consistently. Responsiveness—an instrument's ability to detect changes overtime[19]—was not evident for >90 percent of PREMs. While some of the identified PREMs appear to have been developed for a once-off purpose, and thus exhibiting the ability to detect changes in patient experiences overtime is not a property of significant importance, it was surprising to identify that responsiveness was not evident in most of the CAHPS suite of surveys. Most CAHPS surveys are employed annually on a nationwide scale, such as the HCAHPS, which has been used in this capacity since 2002 in US hospital reimbursement and benchmarking schemes.[38,39] However, only the CAHPS PCMH scored positively for responsiveness. The GPPS PREM also scored positively. This is the UK National General Practitioner Patient Survey which has been undertaken annually since 2007.[40] It is important to note though that this information may be presented outside of the peer-reviewed literature, and consequently, what was captured in this systematic review may be an underrepresentation of all testing undertaken for these measures. The lack of testing for instrument responsiveness is consistent with previous systematic reviews[16,22], both of which identified that responsiveness was not undertaken by any of the PREMs that they assessed. Evidently, testing responsiveness should be prioritized for instruments that are to be utilized on an annual or repeated basis.

The least prevalent property assessed using the COSMIN checklist was measurement error/agreement. Measurement error, in accordance with the revised COSMIN checklist, assesses whether the minimally important change (MIC) (the smallest measured change in participant experience scores that implies practicable importance[41]) is greater than or equal to the smallest detectable change (SDC) in participants scores, or outside of the limits of agreement (LOA) (a technique used when comparing a new measuring technique to what is already practiced[42]). Thus, in the clinical context, the MIC enables researchers to define a threshold of clinical relevancy. That is to say, a score above that threshold (as defined by the MIC) demonstrates that the intervention/program/service was clinically relevant and responsive to improving the patient experience. Given that the patient experience of health care is internationally recognized as a key determinant of health care quality,[32,43] and there is evidence to support the relationship between patient experience data and health care quality,[44,45] the clinical relevancy of improving patient experiences is likely to have implications for resource allocation and decision making in optimizing the quality of health care provided to patients. As such, assessing PREM measurement error/agreement should be undertaken, particularly in instances where PREM scores are being used to inform decision making and funding.

None of the PREMs were tested for all of the revised COSMIN checklist criteria. There are several reasons that this may be the case. For example, criterion validity was only undertaken in roughly 10 percent of the PREMs as some authors recognized that there simply is no gold standard PREM available as a comparator in their given

context.[46,47] Another reason could be inconsistencies in psychometric reporting guidelines and journal guidance regarding what constitutes adequate validity and reliability testing. A previous systematic review[48] examined the quality of survey reporting guidelines. The authors identified that there is generally a lack of validated reporting guidelines for survey instruments. Furthermore, the review highlighted that only a small portion of medical journals, where papers such as those included in this review may be published, provide guidance for the reporting of survey quality.[48] This indicates an area of research generally that warrants greater attention as this is not just a limitation that impacts upon the quality of PREMs, but a wide range of instruments.

4.1 | Limitations

One major limitation of the current study was that grey sources of literature were not considered in the identification of PREMs. Consequently, we may have missed PREMs that otherwise would have fit the inclusion criteria. Furthermore, there were PREMs that we excluded because they had not yet published their supporting validity and reliability results. This was the case for the UK Renal Registry Patient-Reported Experience Measure (UKRR-PREM) who had published their instrument,[49] but were still in the process of developing psychometric evaluation publications at the time that this review was undertaken. However, the purposeful selection of PREMs that were published in peer-reviewed journals was to maximize the quality of the instruments evaluated.

A limitation of the AXIS appraisal tool is that a summative score cannot be derived to interpret the overall quality of the study being assessed[25] (ie, whether a study is deemed poor, moderate, or high quality). However, assessment of risk of bias imposed by a study design is standard practice in the appraisal of studies for systematic reviews.[50] For this study, PREMs were categorized into quartiles according to the proportion of AXIS criteria met, with full details of each PREM assessment provided in Appendix S5 to enable readers to make an informed decision about PREMs that they may use in their own patient experience evaluations and research.

The revised COSMIN checklist also possessed some important limitations. Firstly, the revised version of the COSMIN checklist was used instead of the original checklist[51] as it was more user-friendly to use given the large proportion of PREMs included in this systematic review. Secondly, the parameters of measure for the validity and reliability testing comprising the checklist are very prescriptive. For example, the "structural validity" criteria stated that factors identified through exploratory factor analysis (EFA) had to explain at least 50 percent of the variance.[19] Yet other parameters such as a significant Bartlett's test of sphericity ($P < 0.05$), the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (acceptability typically regarded as >0.6), or factor loading >0.4 (acceptable strength of loading on a factor)[52,53] can also be used to assess the quality of EFA. As such, this limited the authors' ability to assess the reliability and validity of the instruments where tests other than those prescribed in the checklist were undertaken. Thirdly, the checklist fails to attribute rigor to the multidomain design of the included PREMs in measuring the same construct, which may positively impact upon how well the PREM captures a broad array of the attributes of a patient-reported experience.[54] Fourthly, the COSMIN fails to capture the importance of floor and ceiling effects, as well as the percentage of missing data. These were commonly reported statistics among the included PREMs and demonstrate: (a) the ability of the instrument to discern meaningful differences between patients reporting extremes of low and high experience scores; and (b) the burden and feasibility of completing the instrument.[55] Fifthly, the revised COSMIN checklist fails to provide a summative score indicative of whether, overall, a PREM is or is not valid and reliable. Moreover, whether some tests of validity and reliability are more relevant or suitable than other tests to the overall validity and reliability of a PREM remains unknown. Further, it is unclear whether all tests ultimately need to be undertaken in order for a PREM to be labeled as a valid and reliable measure. Thus, in order to assist the reader to make an informed choice in their PREM selection, Appendix S4 ranks the PREMs within their specified contexts, according to the number of "+" scores obtained. Despite these limitations, the COSMIN checklist is currently the most comprehensive psychometric quality criteria for developing outcome measurement instruments and evaluating the method of development for these instruments.[56] Furthermore, the checklist has been applied to other similar systematic reviews[16,22] and was the most appropriate means of systematically measuring the psychometric rigor of the included PREMs.

5 | CONCLUSION

Patient-reported experience measures are internationally recognized instruments for measuring the quality of health care services from the patients perspective. The construct of patient-reported experience appears to still be evolving, and though this systematic review identified PREMs across a range of contexts, PREMs remain largely designed to assess singular events of health care. The key messages of this systematic review are that while the testing of PREM validity and reliability has generally been undertaken in the context of appropriate study designs, there is large variability in both the number and type of validity and reliability testing undertaken for the PREMs identified. As such, it is important that PREM users are aware of the validity and reliability already undertaken for the PREM they have selected, and whether they themselves should undertake more robust testing. Further, the selection of PREMs for research and evaluation purposes should also be considerate of other important selection criteria such as whether a disease/condition or setting-specific measure is more appropriate than a generic measure, and whether a PREM designed in the researcher's country is more appropriate than one designed in a different country, potentially with a different health care system in mind.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: All research was conducted at Griffith University, using University facilities and

equipment. Two authors are employed by the University and two are undertaking PhD degrees. No other disclosures.

ORCID

Claudia Bull https://orcid.org/0000-0003-4064-652X
Martin Downes https://orcid.org/0000-0003-3617-5307

REFERENCES

1. Tremblay D, Roberge D, Berbiche D. Determinants of patient-reported experience of cancer services responsiveness. *BMC Health Serv Res*. 2015;15:425.
2. Schembri S. Experiencing health care service quality: through patients' eyes. *Aust Health Rev*. 2015;39(1):109-116.
3. Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. *Patient*. 2014;7(3):235-241.
4. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *Bja Educ*. 2017;17(4):8.
5. Agency for Healthcare Research and Quality. What is patient experience? 2017; https://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html. Accessed July 18, 2018.
6. Black N, Jenkinson C. Measuring patients' experiences and outcomes. *BMJ*. 2009;339:b2495.
7. Milleson M, Macri J. *Will the Affordable Care Act Move Patient-Centeredness to Centre Stage? Timely Analysis of Immediate Health Policy Issues*. Princeton, NJ: Robert Wood Johnson Foundation and Urban Institute; 2012.
8. Beattie M, Shepherd A, Howieson B. Do the Institute of Medicine's (IOM's) dimensions of quality capture the current meaning of quality in healthcare? - An integrative review. *J Res Nurs*. 2012;18(4):7.
9. European Observatory on Health Systems and Policies. *Pay for Performance in Health Care: Implications for Health Systems Performance and Accountability*. England: World Health Organization (WHO); 2014.
10. Integrated Healthcare Association (IHA). *Value Based Pay for Performance in California: Using Alternative Payment Models to Promote Health Care Quality and Affordability*. California: IHA; 2017.
11. Committee on the Learning Health Care System in America. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: National Academies Press (US): Institute of Medicine; 2013.
12. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. *N Engl J Med*. 2013;368(3):3.
13. Devkaran S. Patient experience is not patient satisfaction: Understanding the fundamental differences. Paper presented at: ISQUA Webinar 2014; online.
14. Shale S. Patient experience as an indicator of clinical quality in emergency care. *Clin Gov*. 2013;18(4):8.
15. Morse JM, Mitcham C, Hupcey JE, Tason MC. Criteria for concept evaluation. *J Adv Nurs*. 1996;24(2):6.
16. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev*. 2015;4:21.
17. Verma R. Overview: What are PROMs and PREMs? NSW: NSW Agency for Clinical Innovation (ACI); n.d.
18. Weldring T, Smith SM. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Serv Insights*. 2013;6:61-68.
19. Gartner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: a systematic review. *PLoS ONE*. 2018;13(2):e0191747.
20. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care*. 2008;14(12):833-838.
21. Terwee CB, Prinsen CA, Ricci Garotti MG, Suman A, de Vet HC, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25(4):767-779.
22. Male L, Noble A, Atkinson J, Marson T. Measuring patient experience: a systematic review to evaluate psychometric properties of patient reported experience measures (PREMs) for emergency care service provision. *Int J Qual Health Care*. 2017;29(3):13.
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
24. Bower WF, Cheung CS, Wong EM, Lee PY, van Hasselt CA. Surgical patient satisfaction in Hong Kong: validation of a new instrument. *Surg Pract*. 2009;13(4):94-101.
25. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*. 2016;6(12):e011458.
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
27. Pallant J. PART FIVE - Statistical techniques to compare groups. In: Pallant J, ed. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*, 4th edn. Berkshire: Open University Press McGraw-Hill Education; 2010:25.
28. Black N, Sanderson C. Day surgery: development of a questionnaire for eliciting patients' experiences. *Qual Health Care*. 1993;2(3):5.
29. Laferriere R. Client satisfaction with home health care nursing. *J Community Health Nurs*. 1993;10(2):67-76.
30. Tian CJ, Tian Y, Zhang L. An evaluation scale of medical services quality based on "patients' experience". *J Huazhong Univ Sci Technolog Med Sci*. 2014;34(2):9.
31. Skudal KE, Garratt AM, Eriksson B, Leinonen T, Simonsen J, Bjertnaes OA. The Nordic Patient Experiences Questionnaire (NORPEQ): cross-national comparison of data quality, internal consistency and validity in four Nordic countries. *BMJ Open*. 2012;2(3):11.
32. Institute of Medicine (IOM). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: IOM; 2001:0309072808.
33. The Commonwealth Fund. What is being done to promote delivery system integration and care coordination? 2018. http://international.commonwealthfund.org/features/integration/. Accessed April 11, 2018.
34. Solomon LS, Hays RD, Zaslavsky AM, Ding L, Cleary PD. Psychometric properties of a group-level Consumer Assessment of Health Plans Study (CAHPS) instrument. *Med Care*. 2005;43(1):53-60.
35. Hays RD, Shaul JA, Williams VS, et al. Psychometric properties of the CAHPS 1.0 survey measures. Consumer Assessment of Health Plans Study. *Med Care*. 1999;37(3 Suppl):MS22-31.
36. Claessen SJ, Francke AL, Sixma HJ, de Veer AJ, Deliens L. Measuring patients' experiences with palliative care: the Consumer Quality Index Palliative Care. *BMJ Support Palliat Care*. 2012;2(4):6.
37. Sjetne IS, Bjertnaes OA, Olsen RV, Iversen HH, Bukholm G. The Generic Short Patient Experiences Questionnaire (GS-PEQ): identification of core items from a survey in Norway. *BMC Health Serv Res*. 2011;11:11.
38. Centers for Medicare and Medicaid Services (CMS), Medicare Learning Network. Hospital Value-Based Purchasing. Internet: U.S. Department of Health & Human Services (HHS); 2017.

39. Agency for Healthcare Research and Quality. About CAHPS. 2018; https://www.ahrq.gov/cahps/about-cahps/index.html. Accessed September 21, 2018.
40. Ipsos MORI. GP Patient Survey. 2018; https://www.gp-patient.co.uk/faq. Accessed September 21, 2018.
41. Coster MC, Nilsdotter A, Brudin L, Bremander A. Minimally important change, measurement error, and responsiveness for the Self-Reported Foot and Ankle Score. *Acta Orthop*. 2017;88(3):300-304.
42. Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. *Br J Anaesth*. 2007;99(3):309-311.
43. NHS Greater Preston CCG. Quality and Clinical Effectiveness. 2018; https://www.greaterprestonccg.nhs.uk/quality-and-clinical-effectiveness/. Accessed May 8, 2018.
44. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev*. 2014;71(5):522-554.
45. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*. 2013;3(1):e001570.
46. Wensing M, Mainz J, Grol R. A standardised instrument for patient evaluations of general practice care in Europe. *Eur J Gen Pract*. 2000;6(3):82-87.
47. Bamm EL, Rosenbaum P, Stratford P. Validation of the measure of processes of care for adults: a measure of client-centred care. *Int J Qual Health Care*. 2010;22(4):302-309.
48. Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: an analysis of published guidance and reporting practices. *PLoS Med*. 2011;8(8):e1001069.
49. Renal Association: UK Renal Registry (UKRR), British Kidney Patient Association, NHS. Patient Experience of Kidney Care: A Report on the Pilot to Test Patient Reported Experience Measures (PREM) in Renal Units in England 2016. UK: UKRR and British Kidney Patient Association; 2016.
50. Higgins JPT, Green S. Assessing risk of bias in included studies. In: Higgins JPT, Altman DG, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Sussex: John Wiley & Sons Ltd; 2008:187-235.
51. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-1179.
52. Chaboyer W, Harbeck E, Bucknall T, et al. Initial psychometric testing and validation of the patient participation in pressure injury prevention scale. *J Adv Nurs*. 2017;73(9):11.
53. Williams B, Onsman A, Brown T. Exploratory factor analysis: a five-step guide for novices. *Australas J Paramed*. 2010;8(3):13.
54. Bollen K, Lennox R. Conventional wisdom on measurement – a structural equation perspective. *Psychol Bull*. 1991;110(2):305-314.
55. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open*. 2015;5(7):e007765.
56. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018.
57. Sawicki GS, Garvey KC, Toomey SL, et al. Development and validation of the adolescent assessment of preparation for transition: a novel patient experience measure. *J Adolesc Health*. 2015;57(3):6.
58. Day C, Michelson D, Hassan I. Child and adolescent service experience (ChASE): measuring service quality and therapeutic process. *Br J Clin Psychol*. 2011;50(4):13.
59. Webster TR, Mantopoulos J, Jackson E, et al. A brief questionnaire for assessing patient healthcare experiences in low-income settings. *Int J Qual Health Care*. 2011;23(3):11.
60. Kertesz SG, Pollio DE, Jones RN, et al. Development of the primary care quality-homeless (PCQ-H) instrument: a practical survey of homeless patients' experiences in primary care. *Med Care*. 2014;52(8):9.
61. Bjertnaes O, Iversen HH, Kjollesdal J. PIPEQ-OS–an instrument for on-site measurements of the experiences of inpatients at psychiatric institutions. *BMC Psychiatry*. 2015;15:9.
62. Bruyneel L, Van Houdt S, Coeckelberghs E, et al. Patient experiences with care across various types of mental health care: Questionnaire development, measurement invariance, and patients' reports. *Int J Methods Psychiatr Res*. 2018;27(1):12.
63. Dinger U, Schauenburg H, Ehrenthal JC, Nicolai J, Mander J, Sammet I. Inpatient and day-clinic experience scale (IDES) – a psychometric evaluation. *Z Psychosom Med Psychother*. 2015;61(4):327-341.
64. Eisen SV, Shaul JA, Clarridge B, Nelson D, Spink J, Cleary PD. Development of a consumer survey for behavioral health services. *Psychiatr Serv*. 1999;50(6):793-798.
65. Eisen SV, Shaul JA, Leff HS, Stringfellow V, Clarridge BR, Cleary PD. Toward a national consumer survey: evaluation of the CABHS and MHSIP instruments. *J Behav Health Serv Res*. 2001;28(3):347-369.
66. Evans J, Rose D, Flach C, et al. VOICE: developing a new measure of service users' perceptions of inpatient care, using a participatory methodology. *J Ment Health*. 2012;21(1):5.
67. Garratt A, Bjorngaard JH, Dahle KA, Bjertnaes OA, Saunes IS, Ruud T. The Psychiatric Out-Patient Experiences Questionnaire (POPEQ): data quality, reliability and validity in patients attending 90 Norwegian clinics. *Nord J Psychiatry*. 2006;60(2):8.
68. Jormfeldt H, Svensson B, Arvidsson B, Hansson L. Dimensions and reliability of a questionnaire for the evaluation of subjective experiences of health among patients in mental health services. *Issues Ment Health Nurs*. 2008;29(1):12.
69. Mavaddat N, Lester HE, Tait L. Development of a patient experience Questionnaire for primary care mental health. *Tidsskr Nor Laegeforen*. 2009;18(2):147-152.
70. Oades LG, Law J, Marshall SL. Development of a consumer constructed scale to evaluate mental health service provision. *J Eval Clin Pract*. 2011;17(6):1102-1107.
71. Olsen RV, Garratt AM, Iversen HH, Bjertnaes OA. Rasch analysis of the Psychiatric Out-Patient Experiences Questionnaire (POPEQ). *BMC Health Serv Res*. 2010;10:9.
72. Schroder A, Larsson BW, Ahlstrom G. Quality in psychiatric care: an instrument evaluating patients' expectations and experiences. *Int J Health Care Qual Assur*. 2007;20(2–3):20.
73. Schroder A, Larsson BW, Ahlstrom G, Lundqvist LO. Psychometric properties of the instrument quality in psychiatric care and descriptions of quality of care among in-patients. *Int J Health Care Qual Assur*. 2010;23(6):17.
74. Fernstrom KM, Shippee ND, Jones AL, Britt HR. Development and validation of a new patient experience tool in patients with serious illness. *BMC Palliat Care*. 2016;15(1):99.
75. Harley C, Adams J, Booth L, Selby P, Brown J, Velikova G. Patient experiences of continuity of cancer care: development of a new medical care questionnaire (MCQ) for oncology outpatients. *Value Health*. 2009;12(8):1180-1186.
76. Iversen HH, Holmboe O, Bjertnæs ØA. The Cancer Patient Experiences Questionnaire (CPEQ): reliability and construct validity following a national survey to assess hospital cancer care from the patient perspective. *BMJ Open*. 2012;2(5):15.
77. van der Veer SN, Jager KJ, Visserman E, et al. Development and validation of the Consumer Quality index instrument to measure the experience and priority of chronic dialysis patients. *Nephrol Dial Transplant*. 2012;27(8):8.
78. Weidmer BA, Cleary PD, Keller S, et al. Development and evaluation of the CAHPS (Consumer Assessment of Healthcare Providers and Systems) survey for in-center hemodialysis patients. *Am J Kidney Dis*. 2014;64(5):753-760.

79. Wood KS, Cronley ML. Then and now: examining how consumer communication and attitudes of direct-to-consumer pharmaceutical advertising have changed in the last decade. *Health Commun*. 2014;29(8):814-825.
80. Bosworth A, Cox M, O'Brien A, et al. Development and validation of a patient reported experience measure (PREM) for patients with rheumatoid arthritis (RA) and other rheumatic conditions. *Curr Rheumatol Rev*. 2015;11(1):1.
81. Grotle M, Garratt A, Lochting I, et al. Development of the rehabilitation patient experiences questionnaire: data quality, reliability and validity in patients with rheumatic diseases. *J Rehabil Med*. 2009;41(7):6.
82. Zuidgeest M, Sixma H, Rademakers J. Measuring patients' experiences with rheumatic care: the consumer quality index rheumatoid arthritis. *Rheumatol Int*. 2009;30(2):9.
83. Haugum M, Iversen HH, Bjertnaes O, Lindahl AK. Patient experiences questionnaire for interdisciplinary treatment for substance dependence (PEQ-ITSD): reliability and validity following a national survey in Norway. *BMC Psychiatry*. 2017;17(1):11.
84. Bobrovitz N, Santana MJ, Ball CG, Kortbeek J, Stelfox HT. The development and testing of a survey to measure patient and family experiences with injury care. *J Trauma Acute Care Surg*. 2012;73(5):1332-1339.
85. Bobrovitz N, Santana MJ, Kline T, Kortbeek J, Stelfox HT. The use of cognitive interviews to revise the Quality of Trauma Care Patient-Reported Experience Measure (QTAC-PREM). *Qual Life Res*. 2015;24(8):1911-1919.
86. Bobrovitz N, Santana MJ, Kline T, et al. Multicenter validation of the quality of trauma care patient-reported experience measure (QTAC-PREM). *J Trauma Acute Care Surg*. 2016;80(1):8.
87. Rattray J, Johnston M, Wildsmith JAW. The intensive care experience: development of the ICE questionnaire. *J Adv Nurs*. 2004;47(1):64-73.
88. Mira JJ, Nuño-Solinís R, Guilabert-Mora M, et al. Development and validation of an instrument for assessing patient experience of chronic Illness care. *Int J Integr Care*. 2016;16(3):13.
89. Homa K, Sabadosa KA, Nelson EC, Rogers WH, Marshall BC. Development and validation of a cystic fibrosis patient and family member experience of care survey. *Qual Manag Health Care*. 2013;22(2):100-116.
90. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring client experiences in maternity care under change: development of a questionnaire based on the WHO Responsiveness model. *PLoS ONE*. 2015;10(2):e0117031.
91. Scheerhagen M, van Stel HF, Tholhuijsen DJC, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ*. 2016;4:21.
92. van der Eijk M, Faber MJ, Ummels I, Aarts JWM, Munneke M, Bloem BR. Patient-centeredness in PD care: development and validation of a patient experience questionnaire. *Parkinsonism Relat Disord*. 2012;18(9):6.
93. Bruyneel L, Tambuyzer E, Coeckelberghs E, et al. New instrument to measure hospital patient experiences in flanders. *Int J Environ Res Public Health*. 2017;14(11):14.
94. Casellas F, Ginard D, Vera I, Torrejon A, GETECCU. Development and testing of a new instrument to measure patient satisfaction with health care in inflammatory bowel disease: the CACHE questionnaire. *Inflamm Bowel Dis*. 2013;19(3):559-568.
95. Cheung CS, Bower WF, Kwok SC, van Hasselt CA. Contributors to surgical in-patient satisfaction–development and reliability of a targeted instrument. *Asian J Surg*. 2009;32(3):143-150.
96. González N, Quintana JM, Bilbao A, et al. Development and validation of an in-patient satisfaction questionnaire. *Int J Qual Health Care*. 2005;17(6):465-472.
97. Jenkinson C, Coulter A, Bruster S. The picker patient experience questionnaire: development and validation using data from in-patient surveys in five countries. *Int J Qual Health Care*. 2002;14(5):6.
98. Jenkinson C, Coulter A, Bruster S, Richards N. The coronary heart disease in-patient experience questionnaire (I-PEQ (CHD)): results from the survey of National Health Service patients. *Qual Life Res*. 2002;11(8):7.
99. Kneebone II, Hull SL, McGurk R, Cropley M. Reliability and validity of the neurorehabilitation experience questionnaire for inpatients. *Neurorehabil Neural Repair*. 2012;26(7):834-841.
100. Labarère J, Fourny M, Jean-Phillippe V, Marin-Pache S, Patrice F. Refinement and validation of a French in-patient experience questionnaire. *Int J Health Care Qual Assur*. 2004;17(1):17-25.
101. Labarere J, Francois P, Auquier P, Robert C, Fourny M. Development of a French inpatient satisfaction questionnaire. *Int J Qual Health Care*. 2001;13(2):99-108.
102. Oltedal S, Bjertnæs Ø, Bjørnsdottìr M, Freil M, Sachs M. The NORPEQ patient experiences questionnaire: data quality, internal consistency and validity following a Norwegian inpatient survey. *Scand J Public Health*. 2007;35(5):8.
103. Pettersen KI, Veenstra M, Guldvog B, Kolstad A. The patient experiences Questionnaire: development, validity and reliability. *Int J Qual Health Care*. 2004;16(6):11.
104. Sullivan PJ, Harris ML, Doyle C, Bell D. Assessment of the validity of the English National Health Service Adult In-Patient Survey for use within individual specialties. *BMJ Qual Saf*. 2013;22(8):690-696.
105. Van Cranenburgh OD, Krol MW, Hendriks MCP, et al. Consumer Quality Index Chronic Skin Disease (CQI-CSD): a new instrument to measure quality of care from the patient's perspective. *Br J Dermatol*. 2015;173(4):1032-1040.
106. Wong EL, Coulter A, Cheung AW, Yam CH, Yeoh EK, Griffiths S. Item generation in the development of an inpatient experience questionnaire: a qualitative study. *BMC Health Serv Res*. 2013;13.
107. Wong EL, Coulter A, Cheung AW, Yam CH, Yeoh EK, Griffiths S. Validation of inpatient experience questionnaire. *Int J Qual Health Care*. 2013;25(4):9.
108. Wong ELY, Coulter A, Hewitson P, et al. Patient experience and satisfaction with inpatient service: development of short form survey instrument measuring the core aspect of inpatient experience. *PLoS ONE*. 2014;10(4):12.
109. Medina-Mirapeix F, del Baño-Aledo ME, Martínez-Payá JJ, Lillo-Navarro MC, Escolar-Reina P. Development and validity of the questionnaire of patients' experiences in postacute outpatient physical therapy settings. *Phys Ther*. 2015;95(5):11.
110. Stubbe JH, Gelsema T, Delnoij DM. The Consumer Quality Index Hip Knee Questionnaire measuring patients' experiences with quality of care after a total hip or knee arthroplasty. *BMC Health Serv Res*. 2007;7:12.
111. Bjertnaes OA, Lyngstad I, Malterud K, Garratt A. The Norwegian EUROPEP questionnaire for patient evaluation of general practice: data quality, reliability and construct validity. *Fam Pract*. 2011;28(3):342-349.
112. Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK – development and psychometric characteristics. *BMC Fam Pract*. 2009;10:10.
113. Campbell JL, Dickens A, Richards SH, Pound P, Greco M, Bower P. Capturing users' experience of UK out-of-hours primary medical care: piloting and psychometric properties of the Out-of-hours Patient Questionnaire. *Tidsskr Nor Laegeforen*. 2007;16(6):462-468.
114. Chanter C, Ashmore S, Mandair S. Improving the patient experience in general practice with the General Practice Assessment Questionnaire (GPAQ). *Qual Prim Care*. 2005;13(4):225-232.
115. Desborough J, Banfield M, Parker R. A tool to evaluate patients' experiences of nursing care in Australian general practice:

development of the Patient Enablement and Satisfaction Survey. *Aust J Prim Health*. 2014;20(2):7.
116. Drain M. Quality improvement in primary care and the importance of patient perceptions. *J Ambul Care Manage*. 2001;24(2):17.
117. Dyer N, Sorra JS, Smith SA, Cleary PD, Hays RD. Psychometric properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS(R)) Clinician and Group Adult Visit Survey. *Med Care*. 2012;50(Suppl):S28-S34.
118. Garratt A, Danielsen K, Forland O, Hunskaar S. The Patient Experiences Questionnaire for Out-of-Hours Care (PEQ-OHC): data quality, reliability, and validity. *Scand J Prim Health Care*. 2010;28(2):7.
119. Greco M, Powell R, Sweeney K. The Improving Practice Questionnaire (IPQ): a practical tool for general practices seeking patient views. *Educ Prim Care*. 2003;14(4):9.
120. Greco M, Sweeney K, Brownlea A, McGovern J. The practice accreditation and improvement survey (PAIS). What patients think. *Aust Fam Physician*. 2001;30(11):5.
121. Holmboe O, Iversen HH, Danielsen K, Bjertnaes O. The Norwegian patient experiences with GP questionnaire (PEQ-GP): reliability and construct validity following a national survey. *BMJ Open*. 2017;7(9):10.
122. Jenkinson C, Coulter A, Gyll R, Lindstrom P, Avner L, Hoglund E. Measuring the experiences of health care for patients with musculoskeletal disorders (MSD): development of the Picker MSD questionnaire. *Scand J Caring Sci*. 2002;16(3):329-333.
123. McInnes DK, Brown JA, Hays RD, et al. Development and evaluation of CAHPS questions to assess the impact of health information technology on patient experiences with ambulatory care. *Med Care*. 2012;50(Suppl):S11-S19.
124. Mead N, Bower P, Roland M. The General Practice Assessment Questionnaire (GPAQ) - development and psychometric characteristics. *BMC Fam Pract*. 2008;9:1.
125. Milano M, Mola E, Collecchia G, et al. Validation of the Italian version of the EUROPEP instrument for patient evaluation of general practice care. *Eur J Gen Pract*. 2007;13(2):3.
126. Paddison C, Elliott M, Parker R, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. *BMJ Qual Saf*. 2012;21(8):634-640.
127. Safran DG, Karp M, Coltin K, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med*. 2006;21(1):9.
128. Scholle SH, Vuong O, Ding L, et al. Development of and field test results for the CAHPS PCMH Survey. *Med Care*. 2012;50(11):S2-S10.
129. Setodji CM, Elliott MN, Abel G, Burt J, Roland M, Campbell J. Evaluating differential item functioning in the English general practice patient survey: comparison of South Asian and White British subgroups. *Med Care*. 2015;53(9):809-817.
130. Steine S, Finset A, Laerum E. A new, brief questionnaire (PEQ) developed in primary health care for measuring patients' experience of interaction, emotion and consultation outcome. *Fam Pract*. 2001;18(4):410-418.
131. Stucky BD, Hays RD, Edelen MO, Gurvey J, Brown JA. Possibilities for shortening the CAHPS clinician and group survey. *Med Care*. 2016;54(1):32-37.
132. Vedsted P, Sokolowski I, Heje HN. Data quality and confirmatory factor analysis of the Danish EUROPEP questionnaire on patient evaluation of general practice. *Scand J Prim Health Care*. 2008;26(3):174-180.
133. Hays RD, Berman LJ, Kanter MH, et al. Evaluating the psychometric properties of the CAHPS Patient-centered Medical Home survey. *Clin Ther*. 2014;36(5):689-696.e681.
134. Chien TW, Wang WC, Lin SB, Lin CY, Guo HR, Su SB. KIDMAP, a web based system for gathering patients' feedback on their doctors. *BMC Med Res Methodol*. 2009;9:10.
135. Garratt A, Bjertnæs ØA, Krogstad U, Gulbrandsen P. The OutPatient Experiences Questionnaire (OPEQ): data quality, reliability, and validity in patients attending 52 Norwegian hospitals. *Tidsskr Nor Laegeforen*. 2005;14:5.
136. Hu Y, Zhang Z, Xie J, Wang G. The Outpatient Experience Questionnaire of comprehensive public hospital in China: development, validity and reliability. *Int J Qual Health Care*. 2017;29(1):7.
137. Loblaw DA, Bezjak A, Bunston T, Loblaw DA, Bezjak A, Bunston T. Development and testing of a visit-specific patient satisfaction questionnaire: the Princess Margaret Hospital Satisfaction With Doctor Questionnaire. *J Clin Oncol*. 1999;17(6):1931-1938.
138. Loblaw DA, Bezjak A, Singh PM, et al. Psychometric refinement of an outpatient, visit-specific satisfaction with doctor questionnaire. *Psychooncology*. 2004;13(4):223-234.
139. Bos N, Sizmur S, Graham C, Van Stel HF. The accident and emergency department questionnaire: a measure for patients' experiences in the accident and emergency department. *BMJ Qual Saf*. 2013;22(2):8.
140. Bos N, Sturms LM, Schrijvers AJ, van Stel HF. The Consumer Quality index (CQ-index) in an accident and emergency department: development and first evaluation. *BMC Health Serv Res*. 2012;12:284.
141. Bos N, Sturms LM, Stellato RK, Schrijvers AJ, van Stel HF. The Consumer Quality Index in an accident and emergency department: internal consistency, validity and discriminative capacity. *Health Expect*. 2015;18(5):13.
142. O'Cathain A, Knowles E, Nicholl J. Measuring patients' experiences and views of the emergency and urgent care system: psychometric testing of the urgent care system questionnaire. *BMJ Qual Saf*. 2011;20(2):7.
143. Keller S, Martin GC, Ewensen CT, Mitton RH. The development and testing of a survey instrument for benchmarking dental plan performance: using insured patients' experiences as a gauge of dental care quality. *J Am Dent Assoc*. 2009;140(2):9.
144. Narayanan A, Greco M. The Dental Practice Questionnaire: a patient feedback tool for improving the quality of dental practices. *Aust Dent J*. 2014;59(3):15.
145. Noest S, Ludt S, Klingenberg A, et al. Involving patients in detecting quality gaps in a fragmented healthcare system: development of a questionnaire for Patients' Experiences Across Health Care Sectors (PEACS). *Int J Qual Health Care*. 2014;26(3):10.
146. Walker KO, Stewart AL, Grumbach K. Development of a survey instrument to measure patient experience of integrated care. *BMC Health Serv Res*. 2016;16(1):11.
147. Benson T, Potts HWW. Short generic patient experience questionnaire: howRwe development and validation. *BMC Health Serv Res*. 2014;14(1):499.
148. Hendriks SH, Rutgers J, van Dijk PR, et al. Validation of the howRu and howRwe questionnaires at the individual patient level. *BMC Health Serv Res*. 2015;15:8.
149. Malott DL, Fulton BR, Rigamonti D, Myers S. Psychometric testing of a measure of patient experience in Saudi Arabia and the United Arab Emirates. *J Surv Stat Methodol*. 2017;5(3):11.
150. Hargraves JL, Hays RD, Cleary PD. Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 adult core survey. *Health Serv Res*. 2003;38(6 Pt 1):1509-1527.
151. Hays RD, Martino S, Brown JA, et al. Evaluation of a care coordination measure for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare survey. *Med Care Res Rev*. 2014;71(2):192-202.
152. Martino SC, Elliott MN, Cleary PD, et al. Psychometric properties of an instrument to assess Medicare beneficiaries' prescription drug plan experiences. *Health Care Financ Rev*. 2009;30(3):41-53.
153. Carle AC, Weech-Maldonado R. Does the Consumer Assessment of Healthcare Providers and Systems Cultural Competence Survey

provide equivalent measurement across English and Spanish versions? *Med Care*. 2012;50(9 Suppl 2):S37-S41.
154. Stern RJ, Fernandez A, Jacobs EA, et al. Advances in measuring culturally competent care: a confirmatory factor analysis of CAHPS-CC in a safety-net population. *Med Care*. 2012;50(9 Suppl 2):S49-S55.
155. Weech-Maldonado R, Carle A, Weidmer B, Hurtado M, Ngo-Metzger Q, Hays RD. The Consumer Assessment of Healthcare Providers and Systems (CAHPS) cultural competence (CC) item set. *Med Care*. 2012;50(9 Suppl 2):S22-S31.
156. Arah OA, ten Asbroek AH, Delnoij DM, et al. Psychometric properties of the Dutch version of the Hospital-level Consumer Assessment of Health Plans Survey instrument. *Health Serv Res*. 2006;41(1):284-301.
157. Dockins J, Abuzahrieh R, Stack M. Arabic translation and adaptation of the hospital consumer assessment of healthcare providers and systems (HCAHPS) patient satisfaction survey instrument. *J Health Hum Serv Adm*. 2015;37(4):518-536.
158. Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. Patterns of unit and item nonresponse in the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):2096-2119.
159. Goldstein E, Farquhar M, Crofton C, Darby C, Garfinkel S. Measuring hospital care from the patients' perspective: an overview of the CAHPS Hospital Survey development process. *Health Serv Res*. 2005;40(6 Pt 2):1977-1995.
160. Keller S, O'Malley AJ, Hays RD, et al. Methods used to streamline the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):21.
161. Levine RE, Fowler FJ Jr, Brown JA. Role of cognitive testing in the development of the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):20.
162. O'Malley AJ, Zaslavsky AM, Hays RD, Hepner KA, Keller S, Cleary PD. Exploratory factor analyses of the CAHPS Hospital Pilot Survey responses across and within medical, surgical, and obstetric services. *Health Serv Res*. 2005;40(6 Pt 2):2078-2095.
163. Squires A, Bruyneel L, Aiken LH, et al. Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries. *Int J Qual Health Care*. 2012;24(5):470-475.
164. Weidmer BA, Brach C, Slaughter ME, Hays RD. Development of items to assess patients' health literacy experiences at hospitals for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Hospital Survey. *Med Care*. 2012;50(9 Suppl 2):S12-S21.
165. Westbrook KW, Babakus E, Grant CC. Measuring patient-perceived hospital service quality: validity and managerial usefulness of HCAHPS scales. *Health Mark Q*. 2014;31(2):97-114.
166. Edward GM, Lemaire LC, Preckel B, et al. Patient Experiences with the Preoperative Assessment Clinic (PEPAC): validation of an instrument to measure patient experiences. *Br J Anaesth*. 2007;99(5):7.
167. Stubbe JH, Brouwer W, Delnoij DMJ. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol*. 2007;7:10.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv Res*. 2019;54:1023-1035. https://doi.org/10.1111/1475-6773.13187