RHEUMATOLOGY ADVANCES IN PRACTICE

Rheumatology Advances in Practice 2021;00:1–10
https://doi.org/10.1093/rap/rkab079
Advance Access Publication 30 October 2021

Systematic review and meta analysis

Patient-Reported Experience Measures in outpatient rheumatology care: a systematic review

Madeleine J. Bryant [1,2,3], Jonathon P. Schubert[1,4], Rachel J. Black[1,2,3] and Catherine L. Hill[1,2,3]

Abstract

Objectives. There is a growing acceptance of the need for routine implementation of patient-reported experience measures (PREMs) in health care. Rheumatology patients, as frequent and long-term users of care, stand to benefit from collection of experience-related data. The aim of this study was to perform a systematic review to identify and critically appraise the development and psychometric validation of PREMs in rheumatology.

Methods. Six databases were searched systematically from inception to 14 December 2020: MEDLINE, EMBASE, PsycINFO, SCOPUS, Cochrane and Google Scholar. We included articles in English that described the use or development of PREMs, with results of psychometric testing, in an adult outpatient rheumatology context. This study is registered with PROSPERO (CRD42021233819). Articles were appraised using the COnsensus Based Standards for the selection of health status Measurement Instruments (COSMIN) (i) Risk of Bias checklist and (ii) criteria for good measurement properties.

Results. The search yielded 3809 publications, and six studies met inclusion criteria. All the included studies on PREM development fulfilled COSMIN standards for 'doubtful' or 'inadequate' quality of instrument development. One study fulfilled a 'sufficient' rating for content validity, and the remainder fulfilled 'inconsistent' ratings. During validity testing, studies fulfilled between one and four of the eight COSMIN checklist criteria for good measurement properties.

Conclusion. Methodological concerns regarding instrument development and validation limit the generalizability of the existing six validated PREMs in use in rheumatology contexts. There is a need for further well-designed studies to validate existing and new PREMs in this area.

Key words: Patient-Reported Experience Measures, instrument, psychometric validation, content validity

CLINICAL SCIENCE

Key messages

- Patient-reported experience measures are gaining recognition and use as indicators of health-care quality.
- Few rheumatology patient-reported experience measures are currently in use, and they vary in instrument development, design and content.
- We highlight the need for greater standardization and rigour of development and validation of patient-reported experience measures.

[1]Adelaide Medical School, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, [2]Rheumatology Unit, The Queen Elizabeth Hospital, Woodville, [3]Rheumatology Unit, The Royal Adelaide Hospital, Adelaide and [4]Department of Gastroenterology, The Queen Elizabeth Hospital, Woodville, South Australia, Australia

Submitted 19 August 2021; accepted 9 October 2021

Correspondence to: Madeleine Bryant, Rheumatology Unit, The Queen Elizabeth Hospital, 28 Woodville Road, Woodville South, SA 5011, Australia. E-mail: madeleine.bryant@sa.gov.au

Introduction

The World Health Organization (WHO) in 2015 signalled the need for a fundamental shift in worldwide health-care funding, management and delivery, towards a people-centred and integrated approach [1]. Likewise, the Australian Commission on Safety and Quality in

Health Care (ACSQHC) mandates that an essential function of the Australian health system is the delivery of safe care that reflects the ideal experience of patients [2], a standard also mirrored in guidance from other international peak bodies, including the National Institute for Health and Care Excellence (NICE, in the UK) and the Institute of Medicine (IOM, in the USA) [3–5].

In line with the pursuit of patient-centred and responsive care is a growing body of evidence to support the routine use of both patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) as indicators of health-care quality and as vital sources of information to improve service delivery [6–8]. PREMs can be defined most simply as surveys that capture the patient perspective. More broadly, they capture data on how care occurred, and evaluate the impact of care delivery and care content on patients [6, 9–11]. PREMs can be completed before and after a specific care encounter, or longitudinally over time, in this way capturing an evaluative purview of care processes. They are considered distinct from both PROMs and patient satisfaction surveys. PROMs typically measure domains such as overall health quality, symptom burden or level of impairment, whereas satisfaction surveys frequently encompass multiple constructs, such as patient expectations and preferences for care, and subjective experiences of how well these were met [8, 11].

Patient-reported experience measures (PREMs) are in use across a wide range of medical and surgical specialties worldwide; a recent systematic review identified 88 individual PREMs implemented across inpatient, ambulatory, primary care and other contexts, with an emphasis on capturing experiences of single events of health care [6]. Another recent report on the system-level impact of routine collection of PREMs proposed that closed-loop feedback of patient-experience data translated to service improvement, behavioural change and positive practices at a broad level [12]. This is congruent with the understanding that people-centred health-care services are those that consciously adopt the perspectives of individuals and communities, and are better positioned to deliver benefits such as increased engagement with care, efficiency and cost gains, in addition to improved equity in uptake of services [1].

Although PREM instruments are widely documented, routine implementation in rheumatology services is not widely practised [6]. Despite this, rheumatology outpatients are well positioned to benefit from integration of experience-related data, given the likelihood of need for long-term care, the high frequency of attendance and the impact of rheumatological diagnoses on quality of life. Rheumatology patients can face potential barriers to care owing to geographical, social and cultural characteristics, such as disparity between rural and metropolitan care provision, fragmentation of care between health jurisdictions, and the challenges of providing care for diverse cultural groups and migrant populations [13]. Previous work on the impact of patient-reported measures on person-centred care demonstrates that patient engagement and empowerment can be enhanced by use of PROMs and PREMs, making barriers to care more surmountable [14–16].

The aim of this study was to identify and critically appraise the development and psychometric validation of PREMs in rheumatology contexts worldwide. The ultimate aim was to determine appropriate instruments for routine use in the rheumatology setting.

Methods

This review was performed in accordance with the preferred reporting items for systemic reviews and meta-analyses (PRISMA) statement [17]. The methods adopted for the search strategy, inclusion criteria and analysis were specified in advance in a protocol registered via PROSPERO International Prospective Register of Systematic Reviews in March 2021 (registration CRD42021233819).

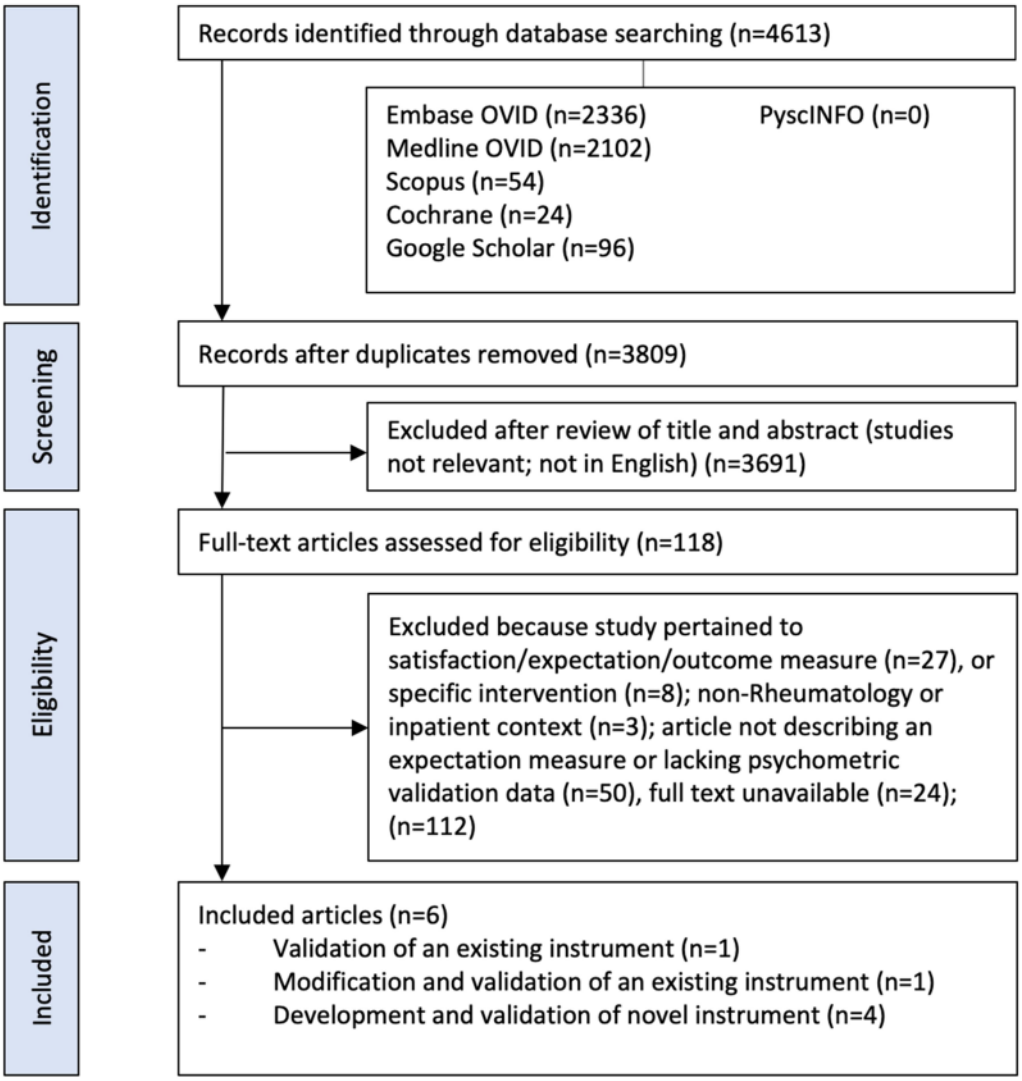
Search strategy and selection criteria

Six databases were searched from inception to 14 December 2020: MEDLINE Ovid (from 1946 to present), EMBASE Ovid (from 1974 to present), PsycINFO Ovid (from 1806 to present), SCOPUS Elsevier, Cochrane Library and Google Scholar. A comprehensive search strategy was adopted with the intention of capturing all relevant articles, given the variable terminology used at present in reference to PREMs in the literature. The full search strategy for each database is available in Supplementary Table S1, available at *Rheumatology Advances in Practice* online. Articles were included if they satisfied the following inclusion criteria: describing the use or development of PREMs or equivalent (including instruments that might be labelled as 'satisfaction' survey, but that measure the patient experience of care), where results of psychometric testing are reported; in an outpatient rheumatology context; published in English or English translation available; and with full-text record available in a peer reviewed journal.

Articles were excluded on the basis of the following criteria: studies describing a satisfaction, expectation or quality of care instrument; studies describing patient outcome measures; studies reporting on patient experience of a specific treatment, intervention or programme; studies in which the PREM psychometric development or validation process was not reported; those reporting a setting other than outpatient rheumatology or adult population (such as inpatient or paediatric); or where the record was available only in abstract form.

After duplicates were removed, a total of 3809 records were identified and screened on the basis of title and abstract, and the full text of 118 records was reviewed by two reviewers (M.J.B. and C.L.H.), with resolution of discrepant votes achieved at a consensus meeting in the presence of a third reviewer (R.J.B.). A large number of articles were excluded after full-text assessment (Fig. 1); most frequently where psychometric

Fig. 1 Preferred reporting items for systemic reviews and meta-analyses (PRISMA) flow diagram



validation data were not reported or the article did not describe PREM use ($n = 50$), or where the article pertained to a measure of satisfaction, expectation or clinical outcome rather than patient experience ($n = 27$). Where only abstracts were available, affiliated articles and pre-publication material were reviewed in order to identify additional records for screening. Six authors were contacted to request supplementary or supporting documents to further the outcome of article eligibility, of whom three responded. A total of six studies were included for analysis (Fig. 1; Table 1).

Data extraction and quality assessment

Data were extracted from each included study on characteristics of subjects participating in instrument development and validation, qualitative study methodology and reported results of psychometric testing per instrument. Quality appraisal of the included articles was performed using the appraisal tool for cross-sectional studies (AXIS) [18]. This instrument allows the rater to consider individual aspects of cross-sectional studies in pursuit of an overall judgement on the study quality.

The COnsensus Based Standards for the selection of health status Measurement Instruments (COSMIN) Risk of Bias checklist was used for critical appraisal of the methodology and results of psychometric testing reported in the included studies [19–21]. This framework was developed through an international collaboration process between expert researchers in health outcome measurement and was selected for quality assessment in this review in the absence of an accepted gold standard equivalent for the appraisal of PREMs. The COSMIN risk of bias checklist addresses the quality of

Table 1 Studies reporting development and/or validation of patient-reported experience measures for outpatient rheumatology use

Author	Year	Instrument	Domain, *n*	Item, *n*	Disease	Recall period	Study design
Beckers *et al.* [28]	2020	CQRA-RA-PREM	8	24	SpA, RA	1 year	Translation of existing CQRA-RA instrument from English to Dutch to English (forward–back translation process) Face validity interviews (participant $n = 16$) Online pilot ($n = 658$) Implementation in practice ($n =$ not reported) Group discussion to identify areas for improvement (details not reported) Action plans formulated and executed (details not reported)
Bosworth *et al.* [26]	2015	CQRA-RA-PREM	8	24	RA	1 year	Item generation: pilot interview (participant $n = 1$), focus group (participant $n = 8$)
		CQRA Rheumatic Conditions-PREM			SS, FM, SLE, gout, PMR, JIA, CBP, OA, inflammatory polyarthritis, SSc		Pilot of draft instrument ($n = 20$) Paper pilot for RA cohort ($n = 524$) Face validity workshops: other rheumatic conditions cohort ($n =$ not reported) Pilot for rheumatic conditions cohort ($n = 110$)
van Campen *et al.* [25]	1998	QUOTE-Rheumatic-Patients	Not stated	32	RA, SpA, OP, OA, LBP	Not stated	Item generation: focus groups ($n =$ not reported) Feasibility testing ($n =$ not reported) Paper pilot ($n = 425$) Adaptation (details not reported)
Guilabert *et al.* [27]	2021	IEXPAC-Rare-Diseases	Not stated	16	APS, EDS, SSc	6 months	Item generation and adaptation of IEXPAC instrument by specialists (details not reported) Face validity testing: patient representatives ($n =$ not reported)
Miedany *et al.* [23]	2014	PREMS	5	32	RA, SpA, PSA	Not stated	Item generation: qualitative interviews (participant $n = 94$) Item reduction: multidisciplinary professional group (details not reported) Paper pilot ($n = 183$) Retest with same subjects ($t = 1$ week)
Zuidgeest *et al.* [24]	2009	CQ-index RA	16	142	RA	1 year	Item generation: focus groups (participant $n = 22$) Face validity testing: combined patient and multidisciplinary professional group (details not reported) Paper pilot ($n = 407$) Item reduction

CBP: chronic back pain; CQRA: commissioning for quality in RA; EDS: Ehlers–Danlos syndrome; IEXPAC: Instrument for the Evaluation of the Experience of Chronic Patients; LBP: low back pain; OP: Osteoporosis; PREM: patient-reported experience measure; QUOTE: Quality of Care Through the Patients' Eyes; *t*: time.

instrument development, and the COSMIN criteria for good measurement properties evaluates instrument validation studies per psychometric measurement property [19–21]. When evaluating overall quality of instrument content validity within the risk of bias checklist, the COSMIN methodology prescribes an appraisal of the domains 'relevance', 'comprehensiveness' and 'comprehensibility', and a judgement regarding whether these domains have been addressed with sufficient, insufficient or inconsistent quality. After completion of the COSMIN Risk of Bias tool, an assessment was made of the level of evidence of the content validity studies using a modified version of the grading of recommendations, assessment, development and evaluations (GRADE) framework [22]. The GRADE methodology, used widely for grading quality of evidence in systematic reviews, was modified by the COSMIN working group for application with PROMs, with the justification that the factors 'imprecision' (confidence intervals) and 'publication bias' are less applicable to this field of study.

Scoring of studies for both the AXIS and COSMIN Risk of Bias tools was performed independently by two reviewers (M.J.B. and J.P.S.). Discrepancies in scoring were resolved by discussion between the two reviewers, and if no consensus could be reached, a third reviewer (C.L.H.) adjudicated the decision. Pooling of results was not performed owing to the heterogeneity of study design and methodology used.

Results

Study characteristics

Of six included studies, four described the development and validation of novel instruments (the CQRA-RA-PREM, PREMs, CQ-Index-RA and the QUOTE-Rheumatic-Patients instruments) [23–26], one the modification and validation of an existing instrument (the IEXPAC-Rare-Diseases instrument) [27], and one the validation of an existing instrument (the CQRA-RA-PREM) [28]. Therefore, a total of five unique PREM instruments were identified by the review. All included studies evaluated patient perception of care within outpatient rheumatology services, in reference to care provided by specialist rheumatologists [23, 26, 28]. Additionally, two instruments specifically included domains pertaining to care provided by non-rheumatologists (e.g. general practitioners, specialty nurses, therapists or surgical specialists) [24, 25], and two reported on the experience of home care services [25, 27]. Within all studies, participants were recruited primarily from tertiary care centres, and two studies reported on additional recruitment from primary care and an insurance company database [24, 25] (Table 1).

Study quality

All six studies satisfied between 13 and 15 of the AXIS criteria [23–28]. The AXIS tool does not provide a numerical scale for assessing the overall quality of a study, thus an overall subjective judgement is required of reviewers [18].

Patient-reported experience measure instrument characteristics

The number of items per PREM ranged from 16 to 142. The recall period ranged from 6 to 12 months [24, 26–28]. Two included studies did not report an intended recall period (Table 1).

Development of PREMs

Three studies lacked data on participant number, age and biological sex of subjects participating in instrument development [24, 25, 27]. Where reported in the remaining three studies, the number of participating subjects was 8, 22 and 94 [23, 24, 26], majority female (72.3–100%), of mean age 51 years (females) and 53 years (males) (Table 2). A range of qualitative study design was evident in the included articles (Table 1).

Patient-reported experience measures content validity assessment

In six studies reporting on PREM content validity assessment, a total of 2568 patients were described, the majority of whom were female, of mean age 41–62.5 years (S.D. 10.1–15.9 years). Median disease duration was 6–8 years (range 0.24–26 years) [23, 26]. Two studies did not report disease duration [24, 25]. Where cohorts were characterized by diagnosis, RA was the most frequently represented

Table 2 Demographic data for subjects participating in instrument development

Author	Item generation method	Participants, *n*	Female, *n* (%)	Age, years
Beckers *et al.* [28]	N/A	N/A	N/A	N/A
Bosworth *et al.* [26]	Patient focus group	8	8 (100)	Median 53 (range 37–71)
van Campen *et al.* [25]	Patient focus group	Not reported	Not reported	Not reported
Guilabert *et al.* [27]	Specialist panel	Not reported	Not reported	Not reported
Miedany *et al.* [23]	Patient interviews	94	68 (72.3%)	Female, mean 51 Male, mean 53
Zuidgeest *et al.* [24]	Patient focus group Professional group	22	Not reported	Not reported

N/A: not assessed.

[$n = 1404$ (54.6%)], followed by SpA [$n = 368$ (14.3%)] (Table 3).

Development and content validity testing involved patients in all six included studies. Three studies described additional consultation with professional groups during the development and validation phases [23, 24, 27].

None of the six included studies reporting on PREM development data satisfied the standards for a COSMIN rating of 'very good' or 'adequate'. Three included studies reporting PREM development fulfilled COSMIN standards for 'doubtful' quality of instrument development [24–26], and two studies for 'inadequate' PREM development [23, 27]. One study fulfilled a 'sufficient' rating [28], and five studies fulfilled 'inconsistent' ratings for overall content validity [23–27]. The level of evidence was assessed by GRADE as moderate quality in five included studies [23–26, 28] and low quality in one study [27] (Table 4).

Methods used to test the psychometric validity were variable among included studies (Table 5). According to COSMIN criteria for good measurement properties, included studies fulfilled between one and three criteria out of a total of eight (Table 4). None fulfilled all prescribed COSMIN criteria. Internal consistency was the criterion most frequently fulfilled, whereas criteria measurement error, hypotheses testing, cross-cultural validity/measurement invariance and responsiveness were not fulfilled by any of the studies.

Discussion

This review demonstrates that only a small number of PREMs are currently in use in rheumatology contexts worldwide, with broad heterogeneity of instrument design and development, delivery and content. With publication of two validated rheumatology-specific instruments in 2020, it is plausible that an awareness of the importance of PREMs in this context is growing, a phenomenon already recognized in the literature regarding uptake of PREMs in general [6].

Lack of reporting of demographic data in existing PREM development studies poses significant shortcomings and might limit the generalizability and utility of these instruments. Guidance suggests that item generation and development require sample sizes approximating 45–50 participants in focus groups and interviews in order to achieve data saturation [29–31]. Only one study reporting demographic data for item development described a sample size within this range ($n = 94$ participants in interviews and focus groups, developing the PREMs instrument) [23], with two studies reporting much smaller cohorts (developing the CQRA-RA-PREM and CQ-Index instruments) [24, 26], and the remainder not reporting these data at all [25, 27]. An essential element of PREM development is the inclusion of members of the target population to ensure sound representation of all those for whom the instrument is intended [21, 32, 33]. However, the majority of participants in both instrument development and validation were female, suggesting that males were under-represented in these processes. Further examples of concerns regarding representation include lack of data on inclusion of different ethnicities or cultural groups, and the disproportionately small cohort of rheumatology patients included in the validation of one instrument (rheumatological diagnoses, $n = 21$ of 261) [27]. These limitations could be overcome by PREM development studies conducted in larger participant samples and with purposive selection of participants to represent different ages, genders and cultural groups.

Methodological concerns arising from instrument development processes were also raised by this review; none of the included studies satisfied standards for a COSMIN rating of 'very good' or 'adequate' instrument development. Likewise, in reference to overall content validity of instruments, only one study fulfilled a 'sufficient' COSMIN rating [28], and none of the studies was appraised as 'high' certainty level of evidence using GRADE methodology. This is consistent with the judgements on instrument development and content validity; per GRADE methodology, the level of evidence is downgraded for inconsistency and limitations in study design. Several plausible explanations for these findings exist, including inadequate qualitative study design, incomplete reporting of sufficient detail of methods to enable affirmative scoring of studies against standards, or the use of standards that are unnecessarily rigorous. It is prudent to note that the COSMIN methodology was developed for use in appraising PROMs rather than PREMs; although significant overlap exists between the two types of instruments, in practice fewer studies on content validity for a given instrument exist for PREMs. An inherent limitation of using the COSMIN methodology to evaluate PREMs is this paucity of data; the present review identified individual PREMs with a single development and content validity study (with the exception of the one instrument validated in two contexts [26, 28]). It is therefore plausible that flaws in study design or reporting are overstated in the judgement on quality, because of the small number of studies. We suggest that there is broad scope for optimizing the methodology adopted during instrument development for PREMs across this field, in addition to further high-quality studies evaluating content validity.

Lastly, this review demonstrates broad variability in psychometric methods used to validate PREMs. During the validation process for new and adapted instruments, all included studies in the present review undertook several components of instrument validation process, but none completed testing of all measurement properties advocated in the COSMIN guidance. Important elements were omitted from validation testing of the majority of instruments; these included, as examples, testing of instrument responsiveness (piloting an instrument at serial time points), measurement invariance (difference between groups by age, gender or language) and measurement error (differences in scores relating to random or systematic error). Other desirable

Table 3 Demographic data for subjects participating in validation of outpatient rheumatology patient-reported experience measures

Author	Administration method	Recruitment method	Diagnosis	Participants, *n*	Female, *n* (%)	Age, years (s.d.)	Disease duration, years (range)
Beckers *et al.* [28]	Online	Online registry	SpA	282	135 (47.9)	Mean 52.7 (12.3)	Mean 8.6 (0.6–66.5)
			RA	376	244 (64.9)	Mean 61.5 (11.9)	Mean 7.7 (0.0–44.0)
Bosworth *et al.* [26]	Paper, postal	Outpatient clinic	RA	524	377 (72%	Median 65 (range 55–80)	Median 8 (3.5–15)
			RA, SS, FM, SLE, gout, PMR, JIA, CBP, OA, inflammatory polyarthritis, SSc	110	69.7%	Median 60 (range 18–84)	Not reported
van Campen *et al.* [25]	Paper, postal	Primary care Patient association	RA, SpA, OP, OA, LBP	425	331 (78)	Mean 62 (14.5), (Range 15–95)	Not reported
Guilabert *et al.* [27]	Online	Patient association	APS, EDS, scleroderma	261 [APS 9 (3.4%), EDS 8 (3%), SSc 4 (1.5%)]	34 (13)	Mean 41 (10.1), (Range 30–90)	Mean 7.8 (s.d. 8)
Miedany *et al.* [23]	Paper, in person	Outpatient clinic	RA, SpA, PSA	183 [RA 97 (53%), SpA 86 (47%)]	140 (76)	Mean 57.8 (15.9)	Median 6 (0.25–26)
Zuidgeest *et al.* [24]	Paper, postal	Insurance company files	RA	407	72.70%	Mean 62.9	Not reported

CBP: chronic back pain; EDS: Ehlers–Danlos syndrome; LBP: low back pain; OP: osteoporosis.

Table 4 Performance per study against consensus-based standards for the selection of health status measurement instruments (COSMIN) criteria for good measurement properties

Author	Content validity (GRADE level of evidence)	Structural validity	Internal consistency	Reliability	Measurement error	Hypotheses testing	Cross-cultural validity/measurement invariance	Criterion validity	Responsiveness
Beckers *et al.* [28]	Sufficient (moderate)	Alternative method: homogeneity +	NT	NT	NT	NT	NT	NT	NT
Bosworth *et al.* [26]	Inconsistent (moderate)	NT +	NT	NT	NT	NT	NT	NT	NT
Van Campen *et al.* [25]	Inconsistent (moderate)	Alternative method: simultaneous component analysis +	NT	NT	NT	NT	NT	NT	NT
Guilabert *et al.* [27]	Inconsistent (low)	Alternative method: principal component analysis +	NT	NT	NT	NT	NT	NT	NT
Miedany *et al.* [23]	Inconsistent (moderate)	+ +	+	NT	NT	NT	NT	+	NT
Zuidgeest *et al.* [24]	Inconsistent (moderate)	Alternative method: Exploratory factor analysis +	Alternate method: Item total correlation	NT	NT	NT	NT	NT	NT

+: sufficient; NT: not tested.

TABLE 5 Reported methods used in psychometric validation of outpatient rheumatology patient-reported experience measures

Author	Property tested	Method employed
Beckers *et al.* [28]	Face validity	Patient focus group (participants, $n = 16$)
	Structural validity	Spearman's coefficient (correlation between average domain score and PROs)
	Divergent validity	Homogeneity coefficient
	Internal consistency	Cronbach's α
	Feasibility	Completion times
	Interpretability	Floor and ceiling effect
Bosworth *et al.* [26]	Face validity	Patient focus group (participants, $n = 8$)
	Internal consistency	Cronbach's α
van Campen *et al.* [25]	Face validity	Focus group (n = not reported)
	Structural validity	Confirmatory factor analysis
	Internal consistency	Cronbach's α
	Feasibility	Comparison of quality impact indices within and between health care services
Guilabert *et al.* [27]	Face validity	Not specified
	Structural validity	Exploratory factor analysis
	Internal consistency	Cronbach's α, Rho coefficient
Miedany *et al.* [23]	Face validity	Patient interviews (participants, $n = 94$)
	Structural validity	Rasch INFIT-OUTFIT, exploratory factor analysis
	Criterion validity	Spearman's coefficient (correlation with PROs)
	Internal consistency	Cronbach's α
	Reproducibility	Test–retest
	Comprehensibility	Not specified
Zuidgeest *et al.* [24]	Face validity	Focus group (participants, $n = 22$)
	Structural validity	Exploratory factor analysis
	Internal consistency	Cronbach's α

PRO: Patient Reported Outcome.

properties for PREMs might be difficult to test given this relatively new area of research. The properties hypothesis testing and criterion/construct validity assume the existence of a gold-standard instrument or measure against which the new tool can be compared. In the present review, only one study cited a series of disease-activity and quality-of-life instruments as standards against which the new instrument would be appraised [23]; none of the studies cited a single gold-standard experience-measure equivalent, which might expound the inability of these studies to satisfy such criteria. Furthermore, the COSMIN checklist guidance does not advocate the use of a summary score for this checklist, nor is it stated explicitly that all tests should be undertaken to deliver a judgement on whether an instrument is altogether valid and reliable. Additionally, it is important to note that the COSMIN methodology was made available in 2018, whereas the majority of articles included in the present review pre-date this publication [23–26], which might account for significant variation in the capacity of PREM development and validation studies to meet the stated COSMIN criteria. These are significant limitations of the COSMIN framework and, subsequently, of the present review, as identified in previous publications on PREMs [6]. For this reason, the results describing performance of instruments against the COSMIN checklist criteria in this review must be interpreted judiciously.

We believe that this is the first systematic review to examine PREMs in rheumatology. We have used a rigorous methodology to identify relevant publications and validated methodology to assess these. However, only a small number of studies met criteria for inclusion, which limits capacity for generalizable conclusions. It is likely that there might be other PREMs in development that were not captured, because a number of studies were available only in abstract form. This finding suggests that further data on PREMs in rheumatology might exist in pre-publication form. Furthermore, the COSMIN methodology is intended for analysis of PROMs rather than specifically for PREMs; however, this remains the best available tool for appraising the psychometric validation of patient-reported instruments.

Conclusion

In this review, we identified six validated PREMs for use with rheumatology outpatients. Heterogeneity of study design makes meta-analysis and transparent comparison between different PREMs difficult. Owing to rapid increases in the interest and implementation of PREMs, this work highlights the need for greater standardization

and rigour of methodological processes for development and validation of PREM instruments. The review also demonstrates that instruments may achieve distribution for use despite not being validated using minimum standardized psychometric methods, meaning that findings arising from such instruments must be interpreted with caution. Specifically, there is a need for further well-designed studies to validate existing and new PREMs in this area. Rheumatology patients stand to benefit greatly from routine application of PREMs and integration of experience-related data in quality-improvement processes, but the integrity of such data is underpinned by the requirement for appropriately validated tools.

Acknowledgements

Library support for this research was provided by Ms Anna Holasek, medical librarian at The Queen Elizabeth Hospital, Adelaide. Dr Marloes Zuidgeest provided a translated version of the CQ-Index RA for analysis. M.J.B. is supported by an Australian Government Research Training Program Scholarship, administered through the University of Adelaide.

Funding: No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Disclosure statement: The authors have declared no conflicts of interest.

Data availability statement

The data underlying this article are available in the article and in its online supplementary material. Further clarification will be provided upon reasonable request to the corresponding author.

Supplementary data

Supplementary data are available at *Rheumatology Advances in Practice* online.

References

1 WHO. Global strategy on people-centred and integrated health services: interim report. Geneva, Switzerland: World Health Organization, 2015.

2 Australian Commission on Safety and Quality in Health Care. Patient-centred care: improving quality and safety through partnerships with patients and consumers. Sydney, Australia: Australian Commission on Safety and Quality in Health Care, 2011.

3 Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. BMJ 2014;348:g2225.

4 Institute of Medicine Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academies Press (US), Copyright 2001 by the National Academy of Sciences. All rights reserved, 2001.

5 National Institute for Health and Care Excellence. Patient experience in adult NHS services: improving the experience of care for people using adult NHS services. London, UK : National Institute for Health and Care Excellence, 2012 (updated 2021).

6 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; 54:1023–35.

7 The King's Fund. Patients' experience of using hospital services: an analysis of trends in inpatient surveys in NHS acute trusts in England, 2005–13. London, UK: The King's Fund, 2015.

8 Manary M, Boulding W, Staelin R, Glickman S. The patient experience and health outcomes. N Engl J Med 2013;368:201–3.

9 Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. Br J Anaesth Educ 2017;17:137–44.

10 Williams K, Sansoni J, Morris D, Grootemaat P, Thompson C. Patient-reported outcome measures: literature review. Sydney, Australia: Australian Commission on Safety and Quality in Health Care, 2016.

11 Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. Patient 2014;7:235–41.

12 De Rosis S, Cerasuolo D, Nuti S. Using patient-reported measures to drive change in healthcare: the experience of the digital, continuous and systematic PREMs observatory in Italy. BMC Health Serv Res 2020;20:315.

13 Morand EF, Leech MT. Successes, challenges and developments in Australian rheumatology. Nat Rev Rheumatol 2015;11:430–6.

14 Wheat H, Horrell J, Valderas JM *et al.* Can practitioners use patient reported measures to enhance person centred coordinated care in practice? A qualitative study. Health Qual Life Outcomes 2018;16:223.

15 Nelson EC, Eftimovska E, Lind C *et al.* Patient reported outcome measures in practice. BMJ 2015;350:g7818.

16 Hvitfeldt H, Carli C, Nelson EC *et al.* Feed forward systems for patient participation and provider support: adoption results from the original US context to Sweden and beyond. Qual Manag Health Care 2009;18: 247–56.

17 Page MJ, McKenzie JE, Bossuyt PM *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

18 Downes M, Brennan M, Williams H, Dean R. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016;6:e011458.

19 Prinsen C, Mokkink L, Bouter L *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147–57.

20 Mokkink L, De Vet H, Prinsen C *et al.* COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1171–9.

21 Terwee C, Prinsen C, Chiarotto A *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27:1159–70.

22 Guyatt G, Oxman G, Vist G *et al.*; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.

23 Miedany YE, Gaafary ME, Youssef S, Ahmed I, Palmer D. The arthritic patients' perspective of measuring treatment efficacy: Patient Reported Experience Measures (PREMs) as a quality tool. Clin Exp Rheumatol 2014;32:547–52.

24 Zuidgeest M, Sixma H, Rademakers J. Measuring patients' experiences with rheumatic care: the consumer quality index rheumatoid arthritis. Rheumatol Int 2009;30: 159–67.

25 Van Campen C, Sixma HJ, Kerssens JJ, Peters L, Rasker JJ. Assessing patients' priorities and perceptions of the quality of health care: the development of the QUOTE-Rheumatic-Patients instrument. Br J Rheumatol 1998;37:362–8.

26 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;11:1–7.

27 Guil abert MM, Martínez-García A, Sala-González M, Solas O, Mira JJ. Results of a Patient Reported Experience Measure (PREM) to measure the rare disease patients and caregivers experience: a Spanish cross-sectional study. Orphanet J Rare Dis 2021;16:67.

28 Beckers E, Webers C, Boonen A *et al.* Validation and implementation of a patient-reported experience measure for patients with rheumatoid arthritis and spondyloarthritis in the Netherlands. Clin Rheumatol 2020;39:2889–97.

29 Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149.

30 Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. Public Opin Q 2007;71:287–311.

31 Pope C, van Royen P, Baker R. Qualitative methods in research on healthcare quality. Qual Saf Health Care 2002;11:148–52.

32 Leidy NK, Vernon M. Perspectives on patient-reported outcomes: content validity and qualitative research in a changing clinical trial environment. Pharmacoeconomics 2008;26:363–70.

33 Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess 2004;16:231–43.